Appendix A Additional related work on other model types.

Fuzzy Systems: Hühn and Hüllermeier (2009); Alcalá-Fdez et al. (2011) build fuzzy rule sets, where boundaries of rules are softer. **CNF:** There are works that study CNF (And of Or's) (Malioutov and Meel 2018; Ghosh and Meel 2019; Ghosh et al. 2022; Yu et al. 2020; Ignatiev et al. 2018) that are able to learn DNF after pre- and post-processing. **Rules for Both Classes:** Both classes are defined by a rule set in works of Bénard et al. (2021) and Ignatiev et al. (2021). **Diverse Rule Sets:** Lakkaraju et al. (2016); Zhang and Gionis (2020); Yang et al. (2021); Yang and van Leeuwen (2022) propose to learn *diverse* sparse rule sets that aim for diversity explicitly (unlike ours). **Linear/Logical Rule Ensembles:** Ensemble methods with base learners of the form of rule set models have been studied (e.g., Akyüz and Birbil 2021; Wang et al. 2021). Rules have also been formed into lists and trees (e.g., Angelino et al. 2018). **Posthoc Model Explanations:** Lu and Ester (2019) propose an active learning algorithm for explaining a learned model with rules.

Appendix B Proofs of bounds on rules

Theorem 1 *(Not too many negatives) If* $R^* \in \arg\min L(R)$*, and* $R^* \neq \emptyset$*, then for any* $r \in R^*$*,* $\text{supp}^{S^-}(r) \leq \frac{N^+}{N} - (C_1 + C_2 + C_3)$.

Proof.

The objective function of the optimal solution R^* is

$$\begin{aligned}
&L(R^*) = \frac{N^+}{N} - \text{supp}^{S^+}(R^*) + \text{supp}^{S^-}(R^*) \\
&+ C_1 n_{\text{condition}}(R^*) + C_2 n_{\text{rule}}(R^*) + C_3 n_{\text{value}}(R^*) \\
&\geq \frac{N^+}{N} - \text{supp}^{S^+}(R^*) + \text{supp}^{S^-}(R^*) + C_1 + C_2 + C_3 \\
&\geq \text{supp}^{S^-}(R^*) + C_1 + C_2 + C_3.
\end{aligned}$$

These inequalities become tight when R^* is one rule with one condition and one value covers the whole positive class and some of the negative class. Let $\emptyset$ denote an empty set where there are no rules and all the data points are classified as negative, so the total number of errors are the number of positive data, denoted as N^+. The objective given $\emptyset$ is

$$L(\emptyset) = \frac{N^+}{N}.$$

Since $R^* \in \arg\min_R L(R)$, $L(R^*) \leq L(\emptyset)$, then

$$\text{supp}^{S^-}(R^*) \leq \frac{N^+}{N} - (C_1 + C_2 + C_3).$$

Since $\mathcal{I}^{S^-}(R^*) = \cup_{r \in R^*} \mathcal{I}_r^{S^-}$, then for any $r \in R^*$, $\text{supp}^{S^-}(r) \leq \text{supp}^{S^-}(R^*)$, thus for any $r \in R^*$,

$$\text{supp}^{S^-}(r) \leq \frac{N^+}{N} - (C_1 + C_2 + C_3).$$

□

Theorem 2 *(Sufficient positive support size) If $R^* \in \arg\min L(R)$, then for any $r \in R^*$ of length $len(r)$ and number of rules $val(r)$, $\text{supp}^{S^+}(r) \geq len(r)C_1 + C_2 + val(r)C_3$.*

Proof.

We prove this theorem by contradiction. Let us assume that R^* has a rule r of length $len(r)$ and number of values $val(r)$ with positive support size lower than $len(r)C_1 + C_2 + val(r)C_3$. Let $R^*_{\setminus r}$ be the rule set with rule r removed from R^*. Then, $R^*_{\setminus r}$ obeys

$$\begin{aligned} L\left(R^*_{\setminus r}\right) =& L(R^*) + \frac{\ell\left(R^*_{\setminus r}\right) - \ell(R^*)}{N} \\ &+ C_1\left(n_{\text{condition}}\left(R^*_{\setminus r}\right) - n_{\text{condition}}(R^*)\right) \\ &+ C_2\left(n_{\text{rule}}\left(R^*_{\setminus r}\right) - n_{\text{rule}}(R^*)\right) \\ &+ C_3\left(n_{\text{value}}\left(R^*_{\setminus r}\right) - n_{\text{value}}(R^*)\right) \\ \leq& L(R^*) + \text{supp}^{S^+}(r) - len(r)C_1 - C_2 - val(r)C_3. \end{aligned}$$

The last inequality is met at equality in the worst case: when all points satisfying the rule r are positive. We would have a contradiction $L\left(R^*_{\setminus r}\right) \leq L(R^*)$ if:

$$L(R^*) + \text{supp}^{S^+}(r) - len(r)C_1 - C_2 - val(r)C_3 \leq L(R^*),$$

i.e., $\text{supp}^{S^+}(r) \leq len(r)C_1 + C_2 + val(r)C_3$. □

Theorem 3 *(Max rule set size) Let $L_{\text{BestSoFar}}[t]$ be the best current objective at iteration t, for rule set R_t. If $R^* \in \arg\min L(R)$, then the number of rules in R^* obeys*

$$n_{rule}(R^*) \leq \frac{L^{[t]}_{\text{BestSoFar}}}{C_1 + C_2 + C_3}. \tag{B1}$$

Proof. Since R^* should be at least as good as the best solution found so far, $L(R^*) \leq L^{[t]}_{\text{BestSoFar}}$, i.e.,

$$\begin{aligned} L^{[t]}_{\text{BestSoFar}} \geq & \frac{\ell\left(R^*\right)}{N} + C_1 n_{\text{condition}}\left(R^*\right) + C_2 n_{\text{rule}}\left(R^*\right) \\ & + C_3 n_{\text{value}}\left(R^*\right) \\ \geq & \ 0 + C_1 n_{\text{conditions}}(R^*) + C_2 n_{\text{rule}}\left(R^*\right) + C_3 n_{\text{value}}\left(R^*\right) \\ \geq & \ (C_1 + C_2 + C_3) n_{\text{rule}}(R^*). \end{aligned}$$

The result directly follows. □

Appendix C Bounds on rules with weighted misclassification errors

Here, we derive bounds for a new objective, where the first term consists of weighted misclassification errors.

$$L_\lambda\left(R\right) = \frac{\ell_\lambda\left(R\right)}{N} + C_1 n_{\text{condition}}\left(R\right) + C_2 n_{\text{rule}}\left(R\right) + C_3 n_{\text{value}}\left(R\right),$$

where $\ell_\lambda\left(R\right) = \lambda \cdot \#\text{positives not covered by } R + \#\text{negatives covered by } R$, with $\lambda \geq 0$.

Theorem 4 *(Not too many negatives) If $R^* \in \arg\min L_\lambda\left(R\right)$, and $R^* \neq \emptyset$, then for any $r \in R^*$, $\text{supp}^{S^-}\left(r\right) \leq \frac{\lambda N^+}{N} - (C_1 + C_2 + C_3)$.*

Proof.

The objective function of the optimal solution R^* is

$$\begin{aligned} & L_\lambda\left(R^*\right) = \lambda\left(\frac{N^+}{N} - \text{supp}^{S^+}\left(R^*\right)\right) + \text{supp}^{S^-}\left(R^*\right) \\ & + C_1 n_{\text{condition}}\left(R^*\right) + C_2 n_{\text{rule}}\left(R^*\right) + C_3 n_{\text{value}}\left(R^*\right) \\ & \geq \lambda\left(\frac{N^+}{N} - \text{supp}^{S^+}\left(R^*\right)\right) + \text{supp}^{S^-}\left(R^*\right) + C_1 + C_2 + C_3 \\ & \geq \text{supp}^{S^-}\left(R^*\right) + C_1 + C_2 + C_3. \end{aligned}$$

These inequalities become tight when R^* is one rule with one condition and one value covers the whole positive class and some of the negative class. Let $\emptyset$ denote an empty set where there are no rules and all the data points are classified as negative, so the total number of errors are the number of positive data, denoted as N^+. The objective given $\emptyset$ is

$$L_\lambda\left(\emptyset\right) = \frac{\lambda N^+}{N}.$$

Since $R^* \in \arg\min_R L_\lambda(R)$, $L_\lambda(R^*) \leq L_\lambda(\emptyset)$, then

$$\text{supp}^{S^-}(R^*) \leq \frac{\lambda N^+}{N} - (C_1 + C_2 + C_3).$$

Since $\mathcal{I}^{S^-}(R^*) = \cup_{r \in R^*} \mathcal{I}_r^{S^-}$, then for any $r \in R^*$, $\text{supp}^{S^-}(r) \leq \text{supp}^{S^-}(R^*)$, thus for any $r \in R^*$,

$$\text{supp}^{S^-}(r) \leq \frac{\lambda N^+}{N} - (C_1 + C_2 + C_3).$$

□

Theorem 5 *(Sufficient positive support size) If $R^* \in \arg\min L_\lambda(R)$, then for any $r \in R^*$ of length $len(r)$ and number of values $val(r)$,* $\text{supp}^{S^+}(r) \geq \frac{len(r)C_1 + C_2 + val(r)C_3}{\lambda}$.

Proof. We prove this theorem by contradiction. Let us assume that R^* has a rule r of length $len(r)$ and number of values $val(r)$ with positive support size lower than $(len(r)C_1 + C_2 + val(r)C_3)/\lambda$. Let $R^*_{\setminus r}$ be the rule set with rule r removed from R^*. Then, $R^*_{\setminus r}$ obeys

$$\begin{aligned}
L_\lambda\left(R^*_{\setminus r}\right) =& L_\lambda(R^*) + \frac{\ell_\lambda\left(R^*_{\setminus r}\right) - \ell_\lambda(R^*)}{N} \\
&+ C_1\left(n_{\text{condition}}\left(R^*_{\setminus r}\right) - n_{\text{condition}}(R^*)\right) \\
&+ C_2\left(n_{\text{rule}}\left(R^*_{\setminus r}\right) - n_{\text{rule}}(R^*)\right) \\
&+ C_3\left(n_{\text{value}}\left(R^*_{\setminus r}\right) - n_{\text{value}}(R^*)\right) \\
\leq& L_\lambda(R^*) + \lambda \text{supp}^{S^+}(r) - len(r)C_1 - C_2 - val(r)C_3.
\end{aligned}$$

The last inequality is met at equality in the worst case: when all points satisfying the rule r are positive. We would have a contradiction $L_\lambda\left(R^*_{\setminus r}\right) \leq L_\lambda(R^*)$ if:

$$L_\lambda(R^*) + \lambda \text{supp}^{S^+}(r) - len(r)C_1 - C_2 - val(r)C_3 \leq L_\lambda(R^*),$$

i.e.,

$$\text{supp}^{S^+}(r) \leq \frac{len(r)C_1 + C_2 + val(r)C_3}{\lambda}.$$

□

Theorem 6 *(Dynamic bound on max size) Let $L_{\lambda\,\text{BestSoFar}}[t]$ be the best current objective at iteration t, for rule set R_t. If $R^* \in \arg\min L_\lambda(R)$, then the number of rules in R^* obeys*

$$n_{rule}(R^*) \leq \frac{L^{[t]}_{\lambda\,\text{BestSoFar}}}{C_1 + C_2 + C_3}. \tag{C2}$$

Proof. Since R^* should be at least as good as the best solution found so far, $L_\lambda(R^*) \leq L^{[t]}_{\lambda\text{BestSoFar}}$, i.e.,

$$\begin{aligned}
&L^{[t]}_{\lambda\text{BestSoFar}} \\
&\geq \frac{\ell_\lambda(R^*)}{N} + C_1 n_{\text{condition}}(R^*) + C_2 n_{\text{rule}}(R^*) \\
&\quad + C_3 n_{\text{value}}(R^*) \\
&\geq 0 + C_1 n_{\text{conditions}}(R^*) + C_2 n_{\text{rule}}(R^*) + C_3 n_{\text{value}}(R^*) \\
&\geq (C_1 + C_2 + C_3) n_{\text{rule}}(R^*).
\end{aligned}$$

The result directly follows. □

Appendix D Bounds on rules for the Rashomon set

Here, we derive bounds for the elements in the ϵ-Rashomon set of R^*, i.e., the elements in $\text{Rash}(R^*; \epsilon, [R]) := \left\{ R \in [R] : L(R) \leq (1+\epsilon) L(R^*) \right\}$, being $R^* \in \arg\min L(R)$ and

$$L(R) = \frac{\ell(R)}{N} + C_1 n_{\text{condition}}(R) + C_2 n_{\text{rule}}(R) + C_3 n_{\text{value}}(R),$$

where $\ell(R) = \#\text{positives not covered by } R + \#\text{negatives covered by } R$.

Theorem 7 *(Not too many negatives) If $\bar{R} \in \text{Rash}(R^*; \epsilon, [R])$, and $\bar{R} \neq \emptyset$, then for any $r \in \bar{R}$, $\text{supp}^{S^-}(r) \leq (1+\epsilon)\frac{N^+}{N} - (C_1 + C_2 + C_3)$.*

Proof. The objective function of $\bar{R} \in \text{Rash}(R^*; \epsilon, [R])$ is

$$\begin{aligned}
&L(\bar{R}) = \frac{N^+}{N} - \text{supp}^{S^+}(\bar{R}) + \text{supp}^{S^-}(\bar{R}) \\
&+ C_1 n_{\text{condition}}(\bar{R}) + C_2 n_{\text{rule}}(\bar{R}) + C_3 n_{\text{value}}(\bar{R}) \\
&\geq \frac{N^+}{N} - \text{supp}^{S^+}(\bar{R}) + \text{supp}^{S^-}(\bar{R}) + C_1 + C_2 + C_3 \\
&\geq \text{supp}^{S^-}(\bar{R}) + C_1 + C_2 + C_3.
\end{aligned}$$

These inequalities become tight when $\bar{R}$ is one rule with one condition and one value covers the whole positive class and some of the negative class. Let $\emptyset$ denote an empty set where there are no rules and all the data points are classified as negative, so the total number of errors are the number of positive data, denoted as N^+. The objective given $\emptyset$ is

$$L(\emptyset) = \frac{N^+}{N}.$$

Since $R^* \in \arg\min_R L(R)$, $L(R^*) \leq L(\emptyset)$, and $L(\bar{R}) \leq (1+\epsilon) L(R^*)$, then

$$\text{supp}^{S^-}(\bar{R}) \leq (1+\epsilon)\frac{N^+}{N} - (C_1 + C_2 + C_3).$$

Since $\mathcal{I}^{S^-}\left(\bar{R}\right) = \cup_{r\in\bar{R}}\mathcal{I}_r^{S^-}$, then for any $r \in \bar{R}$, $\text{supp}^{S^-}(r) \leq \text{supp}^{S^-}\left(\bar{R}\right)$, thus for any $r \in \bar{R}$,

$$\text{supp}^{S^-}(r) \leq (1+\epsilon)\frac{N^+}{N} - (C_1 + C_2 + C_3).$$

□

Theorem 8 *(Sufficient positive support size) If $\bar{R} \in \text{Rash}(R^*; \epsilon, [R])$, then for any $r \in \bar{R}$ of length $len(r)$ and number of values $val(r)$, $\text{supp}^{S^+}(r) \geq len(r)C_1 + C_2 + val(r)C_3$.*

Proof.

We prove this theorem by contradiction. Let us assume that $\bar{R}$ has a rule r of length $len(r)$ and number of values $val(r)$ with positive support size lower than $len(r)C_1 + C_2 + val(r)C_3$. Let $\bar{R}_{\setminus r}$ be the rule set with rule r removed from $\bar{R}$. Then, $\bar{R}_{\setminus r}$ obeys

$$\begin{aligned} L\left(\bar{R}_{\setminus r}\right) =& L\left(\bar{R}\right) + \frac{\ell\left(\bar{R}_{\setminus r}\right) - \ell\left(\bar{R}\right)}{N} \\ &+ C_1\left(n_{\text{condition}}\left(\bar{R}_{\setminus r}\right) - n_{\text{condition}}\left(\bar{R}\right)\right) \\ &+ C_2\left(n_{\text{rule}}\left(\bar{R}_{\setminus r}\right) - n_{\text{rule}}\left(\bar{R}\right)\right) \\ &+ C_3\left(n_{\text{value}}\left(\bar{R}_{\setminus r}\right) - n_{\text{value}}\left(\bar{R}\right)\right) \\ \leq& (1+\epsilon)L(R^*) + \text{supp}^{S^+}(r) - len(r)C_1 - C_2 - val(r)C_3. \end{aligned}$$

The last inequality is met at equality in the worst case: when all points satisfying the rule r are positive. We would have a contradiction $L\left(\bar{R}_{\setminus r}\right) \leq (1+\epsilon)L(R^*)$ if:

$$(1+\epsilon)L(R^*) + \text{supp}^{S^+}(r) - len(r)C_1 - C_2 - val(r)C_3 \leq (1+\epsilon)L(R^*),$$

i.e., $\text{supp}^{S^+}(r) \leq len(r)C_1 + C_2 + val(r)C_3$. □

Theorem 9 *(Dynamic bound on max size) Let $L_{\text{BestSoFar}}[t]$ be the best current objective at iteration t, for rule set R_t. If $\bar{R} \in \text{Rash}(R^*; \epsilon, [R])$, then the number of rules in $\bar{R}$ obeys*

$$n_{rule}(\bar{R}) \leq \frac{(1+\epsilon)\,L^{[t]}_{\text{BestSoFar}}}{C_1 + C_2 + C_3}. \tag{D3}$$

Proof. Since $\bar{R}$ should be at least as good as the best solution found so far, weighted by $(1+\epsilon)$, $L(\bar{R}) \leq (1+\epsilon)L^{[t]}_{\text{BestSoFar}}$, i.e.,

$$\begin{aligned} &(1+\epsilon)L^{[t]}_{\text{BestSoFar}} \\ &\quad\geq \frac{\ell\left(\bar{R}\right)}{N} + C_1 n_{\text{condition}}\left(\bar{R}\right) + C_2 n_{\text{rule}}\left(\bar{R}\right) + C_3 n_{\text{value}}\left(\bar{R}\right) \end{aligned}$$

$$\geq 0 + C_1 n_{\text{conditions}}(\bar{R}) + C_2 n_{\text{rule}}\left(\bar{R}\right) + C_3 n_{\text{value}}\left(\bar{R}\right)$$
$$\geq (C_1 + C_2 + C_3) n_{\text{rule}}(\bar{R}).$$

The result directly follows. □

Appendix E Results for robustness

We illustrate the robustness of FastSRS to both noisy labels and features. For that, we have designed a synthetic data set with $N = 10000$ observations and $J = 200$ randomly generated binary features, including negations. The true sparse rule set R^* consists of 5 rules of maximum length 3, where columns are randomly selected without replacement, to constitute the conditions. Note that, in this setting, $n_{\text{condition}}\left(R^*\right) = n_{\text{value}}\left(R^*\right)$. The labels $\{Y_n\}_{n=1}^{10000}$ are assigned accordingly, ensuring that the positives examples satisfy at least one rule in R^*. The obtained class distribution was $80\% - 20\%$. 5-fold cross-validation was used to split the sample into different training-test subsets.

To test robustness with respect to label noise, we randomly flip a certain percentage of Y_n's, chosen from $\{1\%, 5\%, 10\%, 15\%, 20\%, 25\%\}$. We will indicate the original training sets with 0%. The test subsets were not corrupted. We report the performance of FastSRS in Figure E1. The training accuracy decreases as a function of the mislabeling percentage. However, test accuracy is very robust since the test data does not contain noise. The test accuracy is only slightly damaged, dropping only 2% even when 25% of the training labels are flipped. The result shows that FastSRS is robust to random noise in the training data label. This is because the predictions are made by rules, which cover an area instead of one instance. Thus, even if some instances have errors, they are dominated by nearby instances with no error. Therefore, rule-based models are generally more robust and less likely to overfit, especially when the model is sparse, corresponding to rules with larger support.

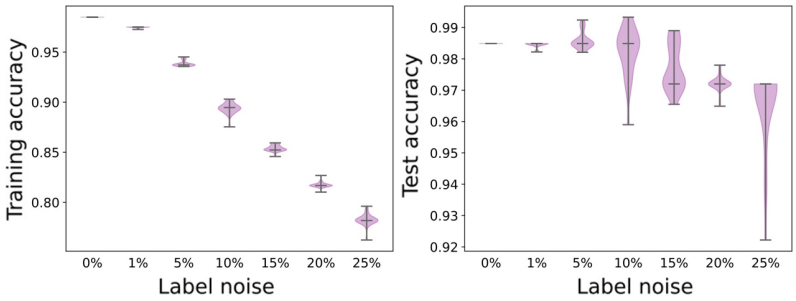

Fig. E1 Robustness of FastSRS to noisy labels

Similar results were obtained in Figure E2 when testing the robustness of our approach with respect to feature noise. In this case, we flipped the features values of the training subsets by different percentages $\{1\%, 5\%, 10\%, 15\%, 20\%, 25\%\}$, and transferred the flips to the negations of the features. While the training accuracy decreases gradually as a function of feature noise, the test accuracy drops less than

1%, even with up to 25% of noisy feature values. More results can be found in Figures E3 and E4 for the robustness of FastSRS to noisy labels and features, respectively.

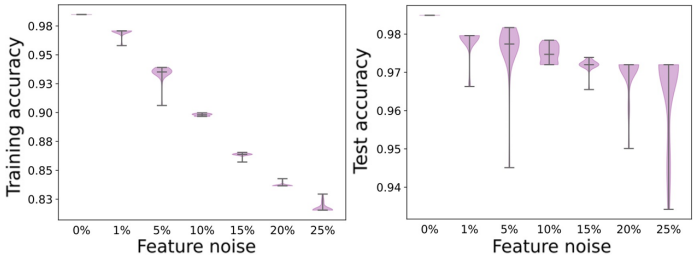

Fig. E2 Robustness of FastSRS to noisy features

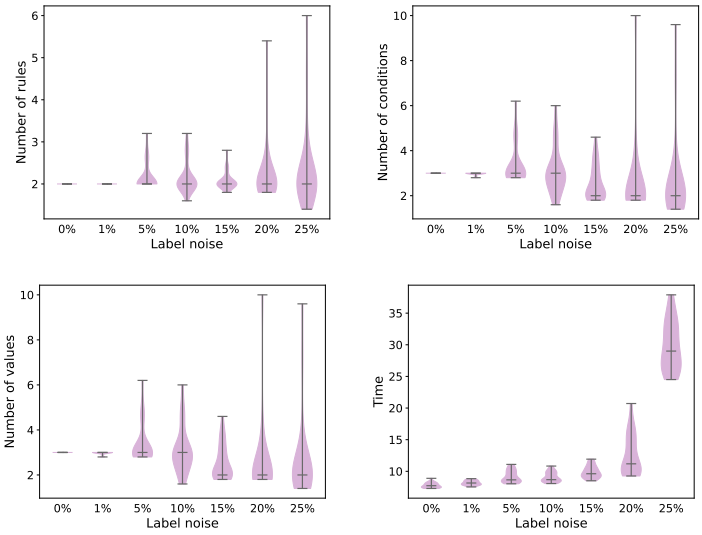

Fig. E3 Additional results of robustness of FastSRS to noisy labels.

Appendix F Sources for the datasets used

The datasets used in the experiments together with their sources are `heart` (Janosi et al. 1988), `diabetes` (OpenML 2014), `compas` (ProPublica 2016), `fico` (OpenML 2026), `recidivism` (National Institute of Justice 2021), `invehicle` (Wang et al. 2020) and `adult` (Kohavi and Becker 1996).

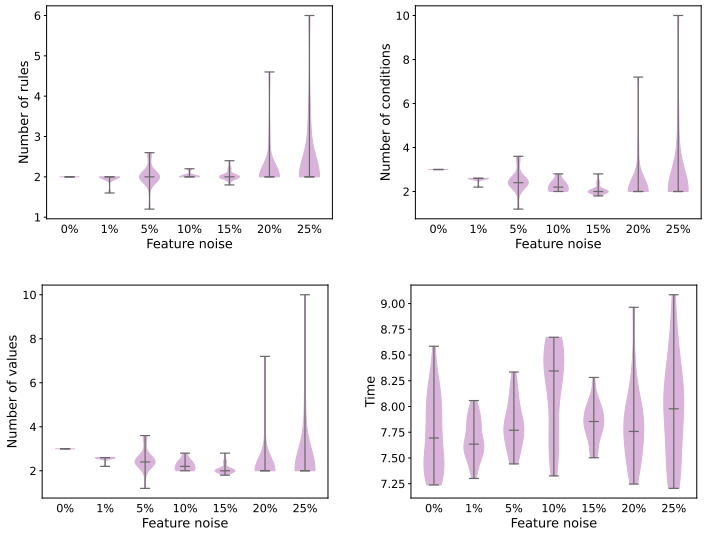

Fig. E4 Additional results of robustness of FastSRS to noisy features.

Appendix G Results for comparison against benchmarks for rule sets

Baseline methods

- BRCG (Dash et al. 2018) with (lambda0, lambda1) $\in \{(x, y) : x, y \in \text{linspace}(0.001, 0.01, 5)\}$, and FeatureBinarizer with negations = False.
- BRS (Wang et al. 2017) with $N_{\text{rules}} = 2000$, $N_{\text{iter}} = 500$, $N_{\text{chain}} = 2$, threspossupp = 5, maxlen = 3, method = randomForest, $\alpha_+ = \alpha_- \in \{100, 1000, 10000\}$, $\beta_+ = \beta_- = 1$, α_l and β_l default terms.
- MRS (Wang 2018) with $N_{\text{rules}} = 5000$, $N_{\text{iter}} = 500$, $N_{\text{chain}} = 1$, threspossupp = 5, maxlen = 3, method = randomForest, $\alpha_+ = \alpha_- = 100$, $\beta_+ = \beta_- = 1$, $\alpha_l = \beta_l = 1$ and $\beta_M = \beta_L \in \{1, 10, 100, 1000, 10000\}$.
- LIBRE (Mita et al. 2020) with default top_K, $n\_estimators \in \{5, 20, 50\}$, $n\_features = 5$, $search\_heuristic \in \{H_1, H_2\}$ and $\alpha \in \{0.5, 0.7, 0.9\}$. LIBRE3 is the same configuration but restricting top_K to 3.
- RIPPER (Cohen 1995) with $k \in \{1, 2\}$ and $prune\_size \in \{0.1, 0.25, 0.33, 0.5\}$

Extra Experimental Results

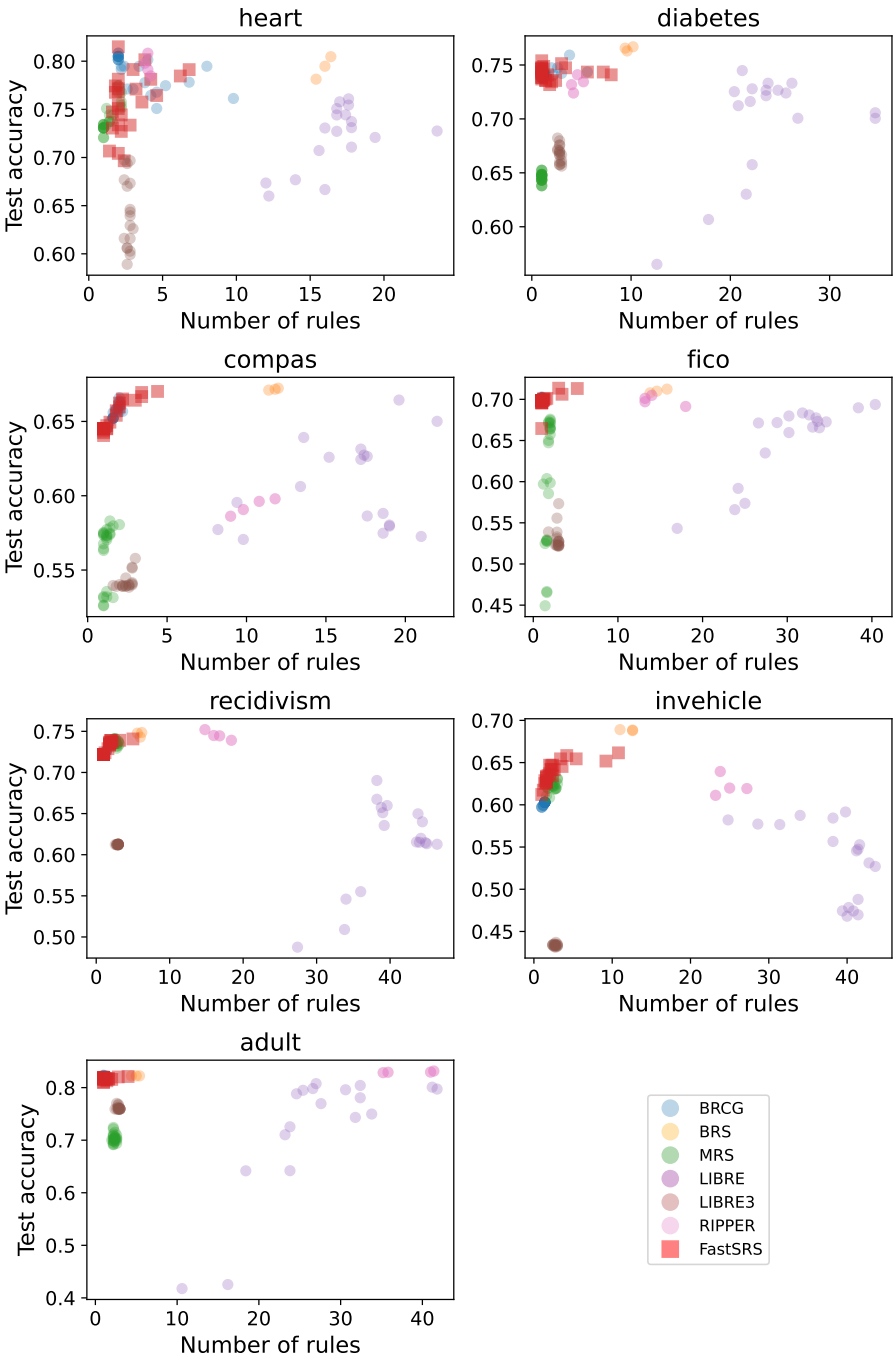

Fig. G5 Comparison against benchmarks in terms of test accuracy and number of rules.

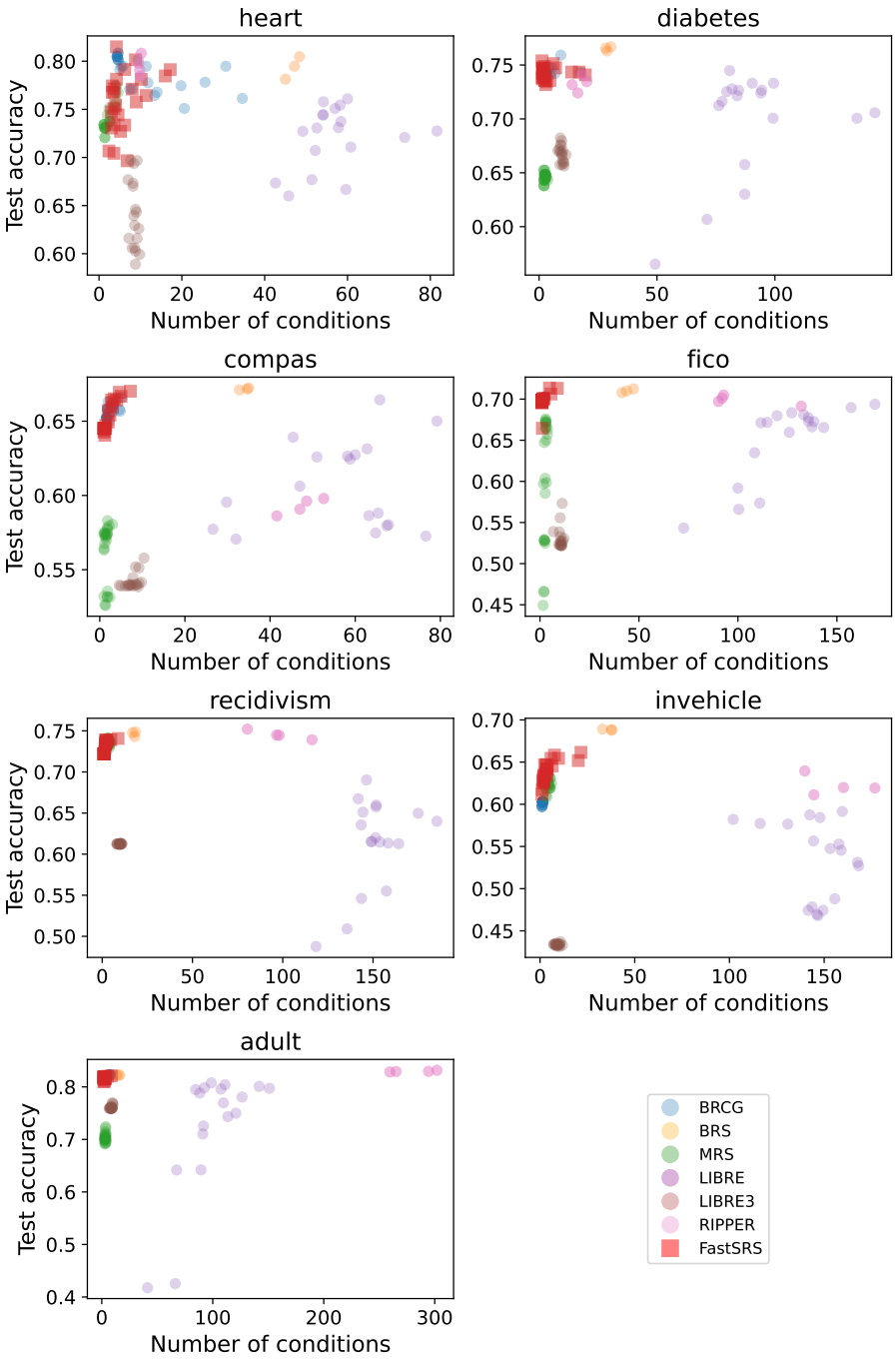

Fig. G6 Comparison against benchmarks in terms of test accuracy and number of conditions.

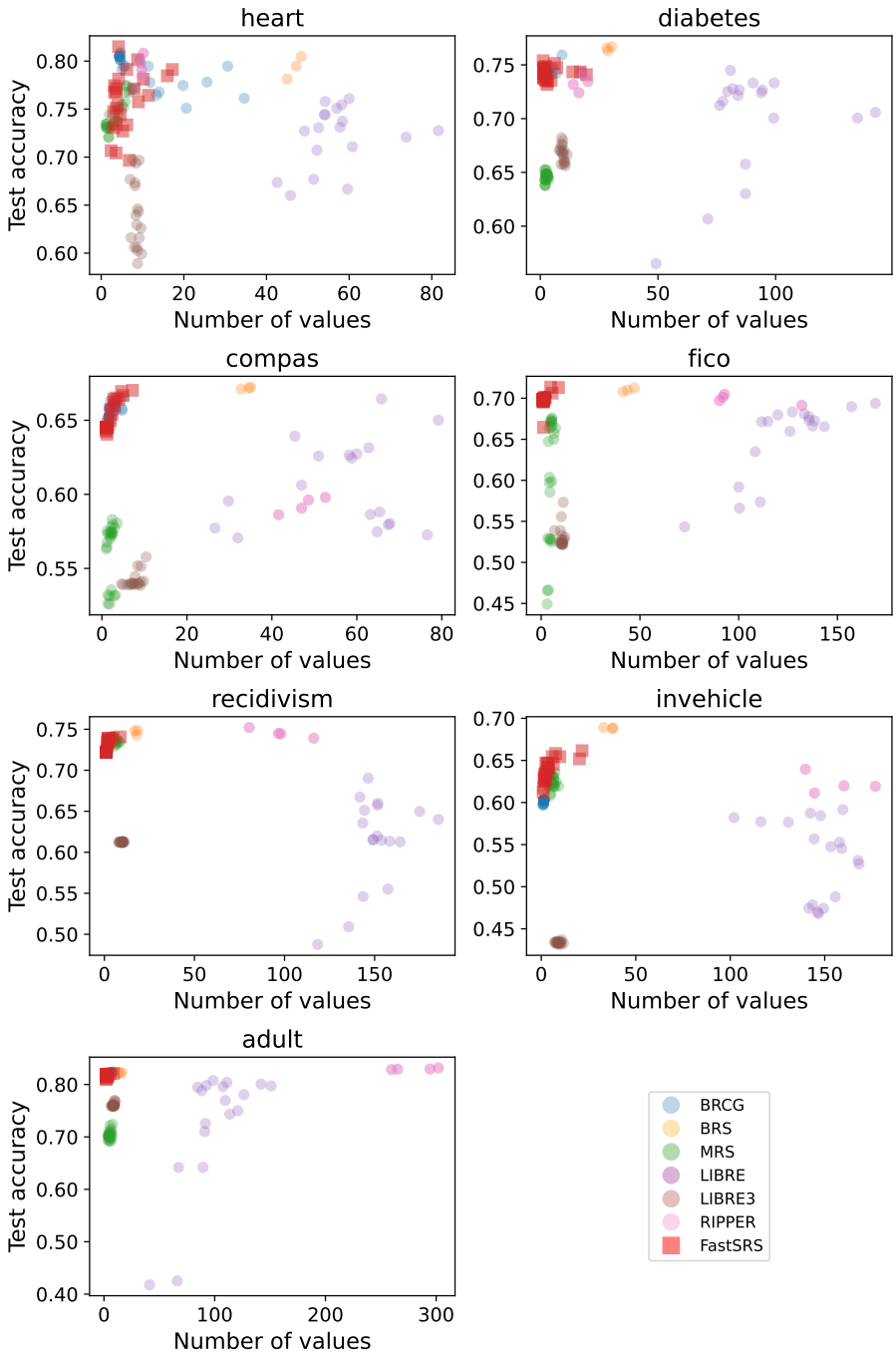

Fig. G7 Comparison against benchmarks in terms of test accuracy and number of values.

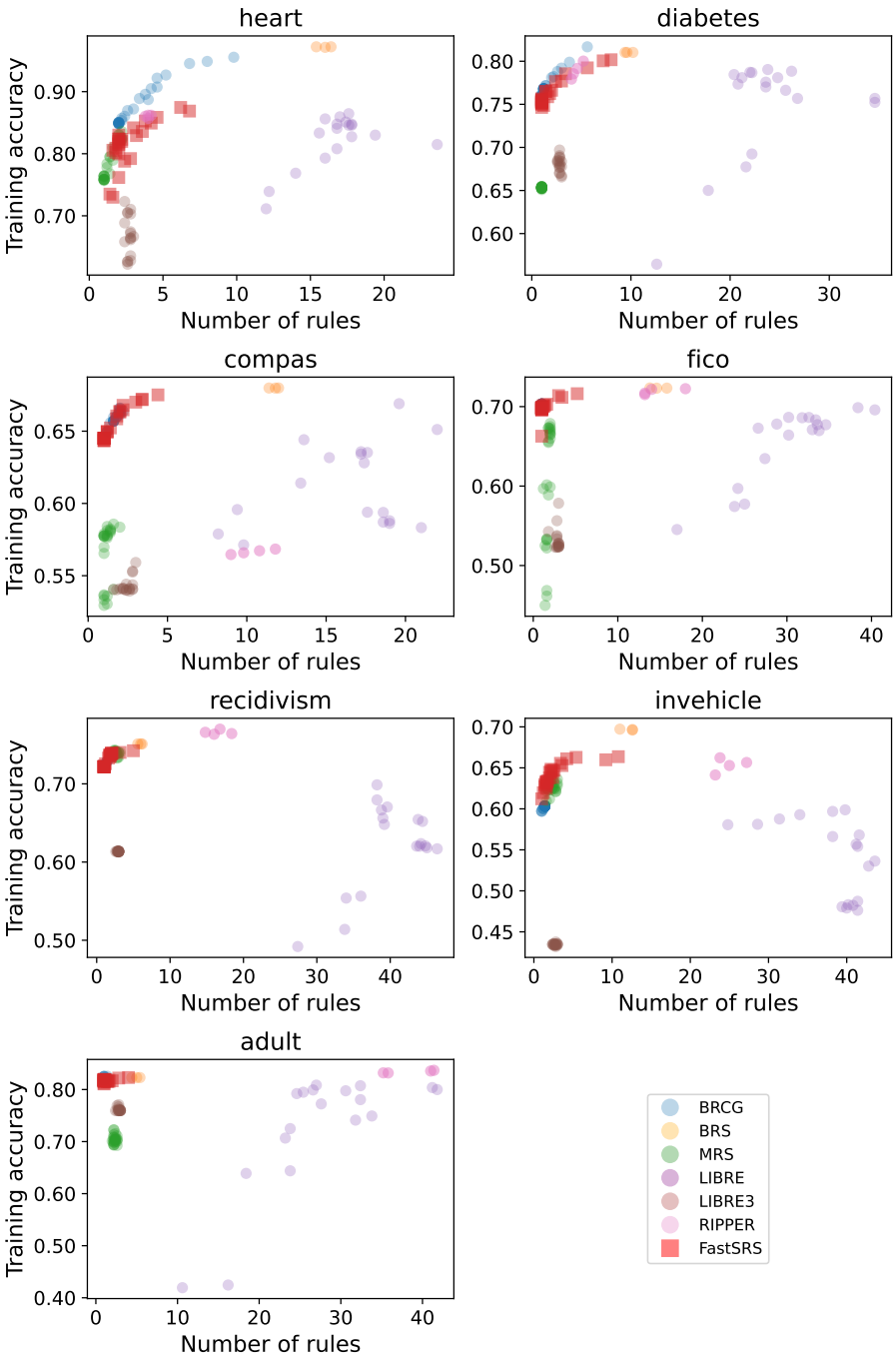

Fig. G8 Comparison against benchmarks in terms of train accuracy and number of rules.

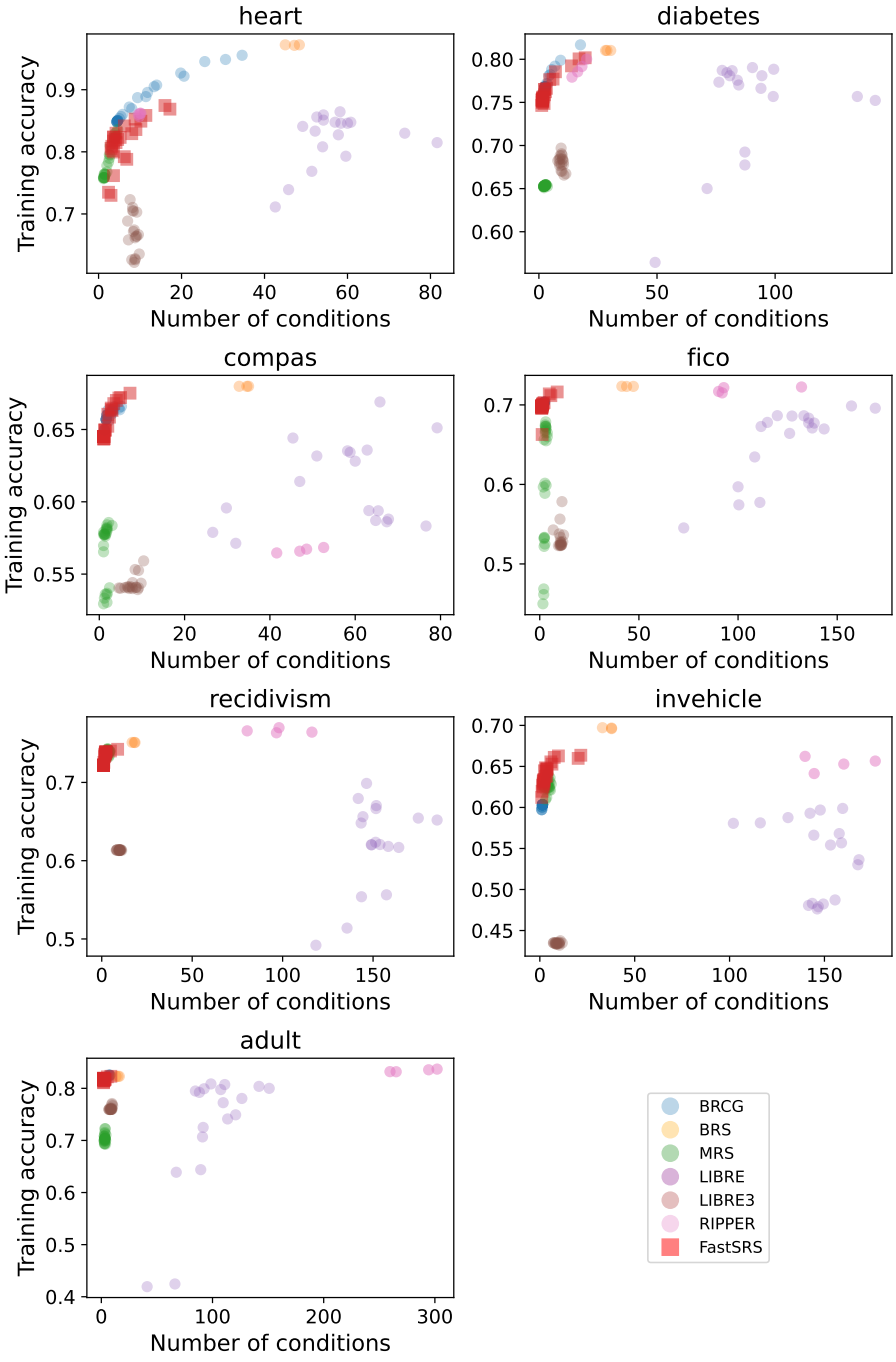

Fig. G9 Comparison against benchmarks in terms of train accuracy and number of conditions.

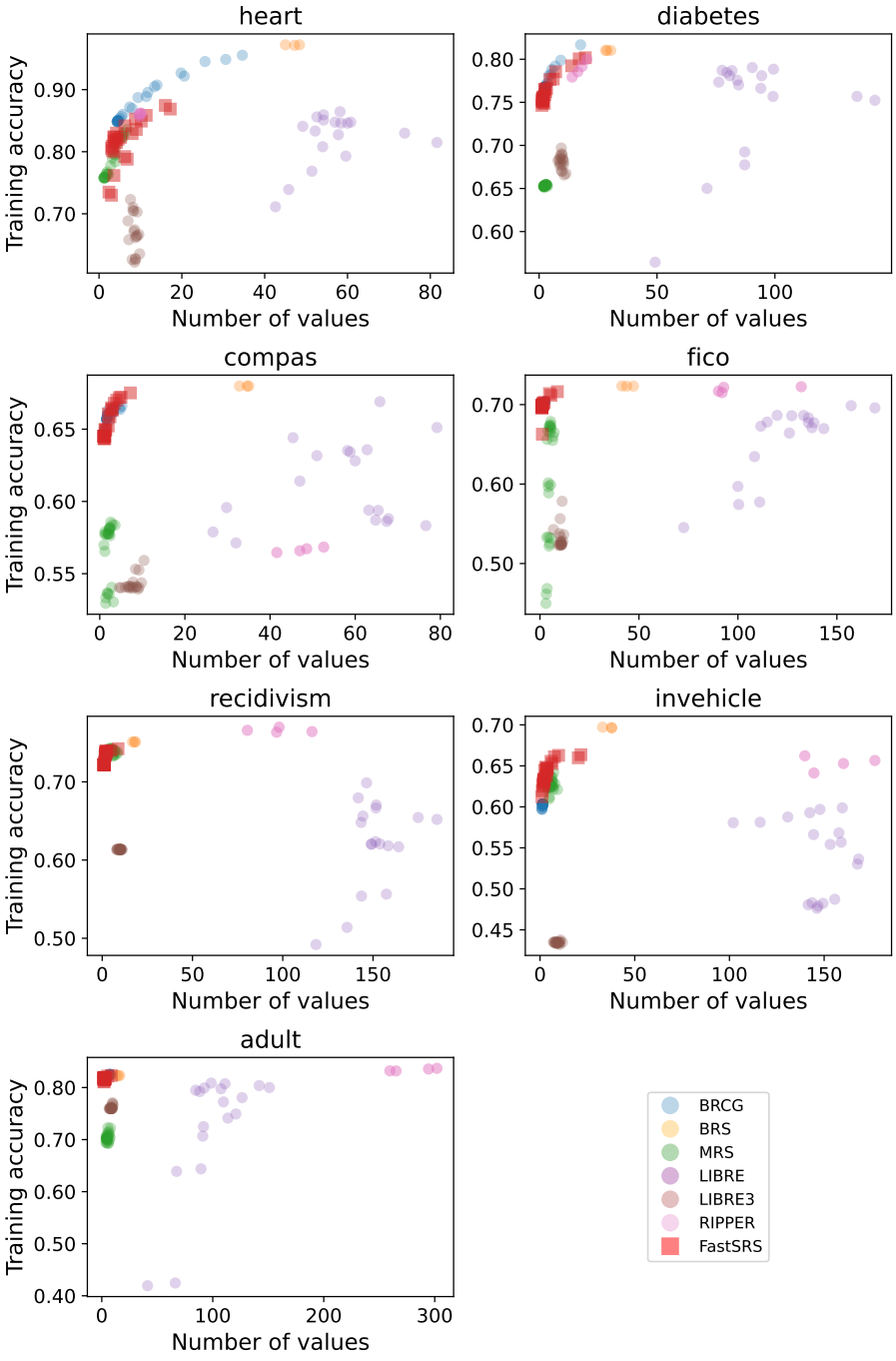

Fig. G10 Comparison against benchmarks in terms of train accuracy and number of values.

Appendix H Approximation quality to the global optimum and stability across runs of FastSRS

First, to assess how close FastSRS's best solution is to the global optimum, we construct a synthetic dataset with $J = 8$ binary features (16 with negations), $N = 1{,}000$ examples, and 5% label noise. The ground truth consists of 3 rules of maximum length 2. We exhaustively enumerate all 275,584,033 possible rule sets (up to 5 rules, each of length at most 2) and compute the true global optimum $R^* = \arg\min L(R)$ for each of 25 sparsity parameter configurations (C_2 ranging from 0.0005 to 0.025, with $C_1 = C_2/4$ and $C_3 = C_1/3$).

FastSRS finds the exact global optimum predictions in 19 out of 25 configurations (76%), with a mean optimality gap of 1.89% across all configurations. In the 6 configurations where FastSRS does not match the true optimum (gap up to 17.1%), the algorithm converges to a rule set with fewer rules than the true optimum.

Second, to assess stability across runs, we ran FastSRS with different random seeds on the same configurations. The results are consistent across seeds: FastSRS reliably finds the same best model in each configuration, demonstrating stable convergence behavior. See Table H1.

Table H1 Convergence analysis across 5 seeds, 5 folds, and 25 (C_1, C_2, C_3) configurations. *Obj std* denotes the standard deviation in the objective value across seeds. *Acc std* denotes the standard deviation in accuracy across seeds. *Agreement* denotes mean and minimum pairwise prediction agreement across seeds. *Exact match* denotes the percentage of seed pairs producing identical prediction vectors.

Dataset	Obj std	Acc std	Agreement		Exact match
			Mean	Min	
heart	0.000000	0.0000	100.00%	100.00%	100.0%
compas	0.000000	0.0000	100.00%	100.00%	100.0%
adult	0.000000	0.0000	100.00%	100.00%	100.0%
diabetes	0.000000	0.0000	100.00%	100.00%	100.0%
fico	0.000044	0.0030	99.97%	91.10%	99.7%
recidivism	0.000000	0.0000	100.00%	100.00%	100.0%
invehicle	0.000000	0.0000	100.00%	100.00%	100.0%

Appendix I More details of the example

Table I2 has an explanation of features in the Janosi et al. (1988) dataset.

Attributes		
Feature	Type	Description
age	Integer	age in years
exang	Binary	1 if exercise induced angina
ca	Integer	number of major vessels (0-3) colored by flourosopy
cp	Categorical	chest pain type (1: typical angina, 2: atypical angina, 3: non-anginal pain, 4: asymptomatic)
fbs	Binary	1 if fasting blood sugar 120 mg/dl
chol	Numerical	serum cholestoral in mg/dl
oldpeak	Numerical	ST depression induced by exercise relative to rest
restecg	Binary	resting electrocardiographic results (0: normal, 1: having ST-T wave abnormality, 2: showing probable or definite left ventricular hypertrophy by Estes' criteria)
sex	Binary	1 if sex is male
slope	Categorical	the slope of the peak exercise ST segment (1: upsloping, 2: flat, 3: downsloping)
thal	Categorical	thalassemia (3: normal, 6: fixed defect, 7: reversable defect)
thalach	Numerical	maximum heart rate achieved
trestbps	Numerical	peak exercise blood pressure

Table I2 Features in the Janosi et al. (1988) dataset.

Appendix J Results for comparison against benchmark for Rashomon sets

Baseline method

The only existing method for calculating Rashomon sets of rule sets is ERS (Ciaperoni et al. 2024). We used parameter values $\lambda \in \{0.1, 0.01, 0.15, 0.05\}$, $\theta \in \{0.7, 0.9, 1.1\}$, $\epsilon = 8$, $|\mathcal{U}| \in \{30, 50\}$, n_rule_sets $\in \{5, 10, 15\}$, and the default options for the remaining parameters.

Extra Experimental Results

Figures J11 through J14 show test accuracy vs sparsity. Figures J15 and J16 show test accuracy vs diversity disagreement.

Figures J17 through J20 provide training accuracy vs sparsity. Figures J21 and J22 show training accuracy vs diversity disagreement.

Computation time across the rule sets in the Rashomon set is in Figures J23 and J24. Test/train accuracy vs computation is in Figures J25 through J28.

Model size (measured in multiple ways) vs computation time is in Figures J29 through J32.

Figure J33 shows Rashomon set sizes as a function of the Rashomon parameter.

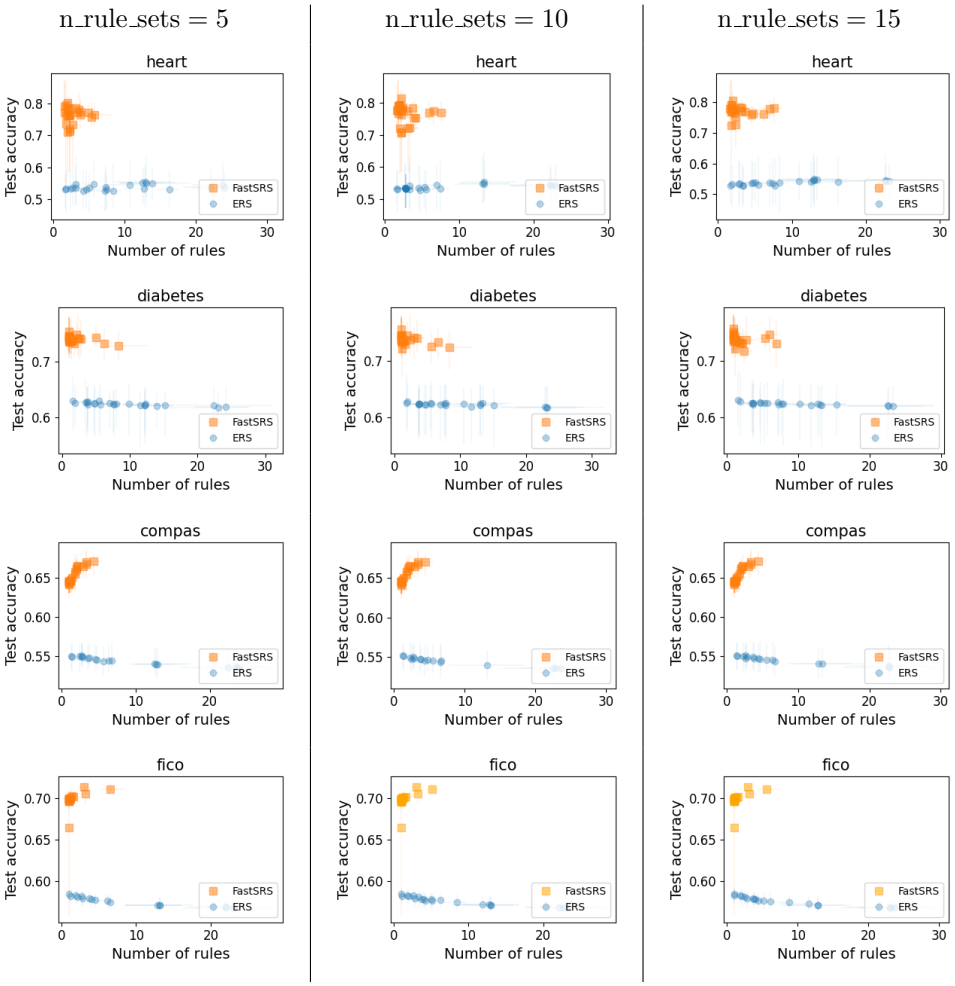

Fig. J11 Average accuracy test versus average number of rules across the rule sets in the Rashomon set for all configuration tested (1/2).

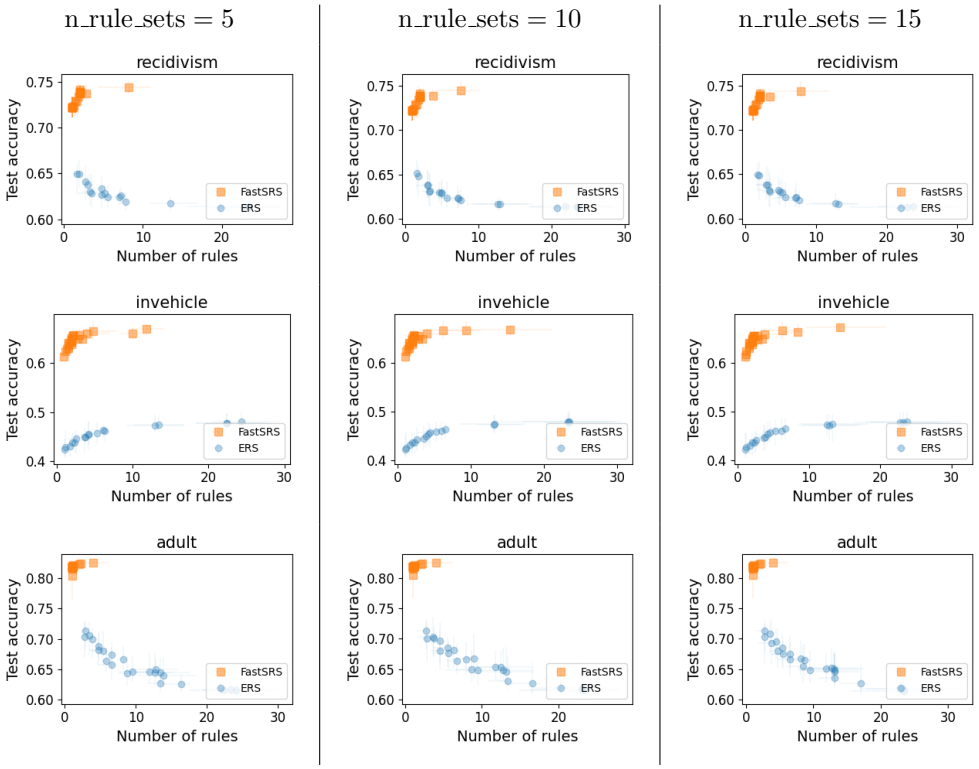

Fig. J12 Average accuracy test versus average number of rules across the rule sets in the Rashomon set for all configuration tested (2/2).

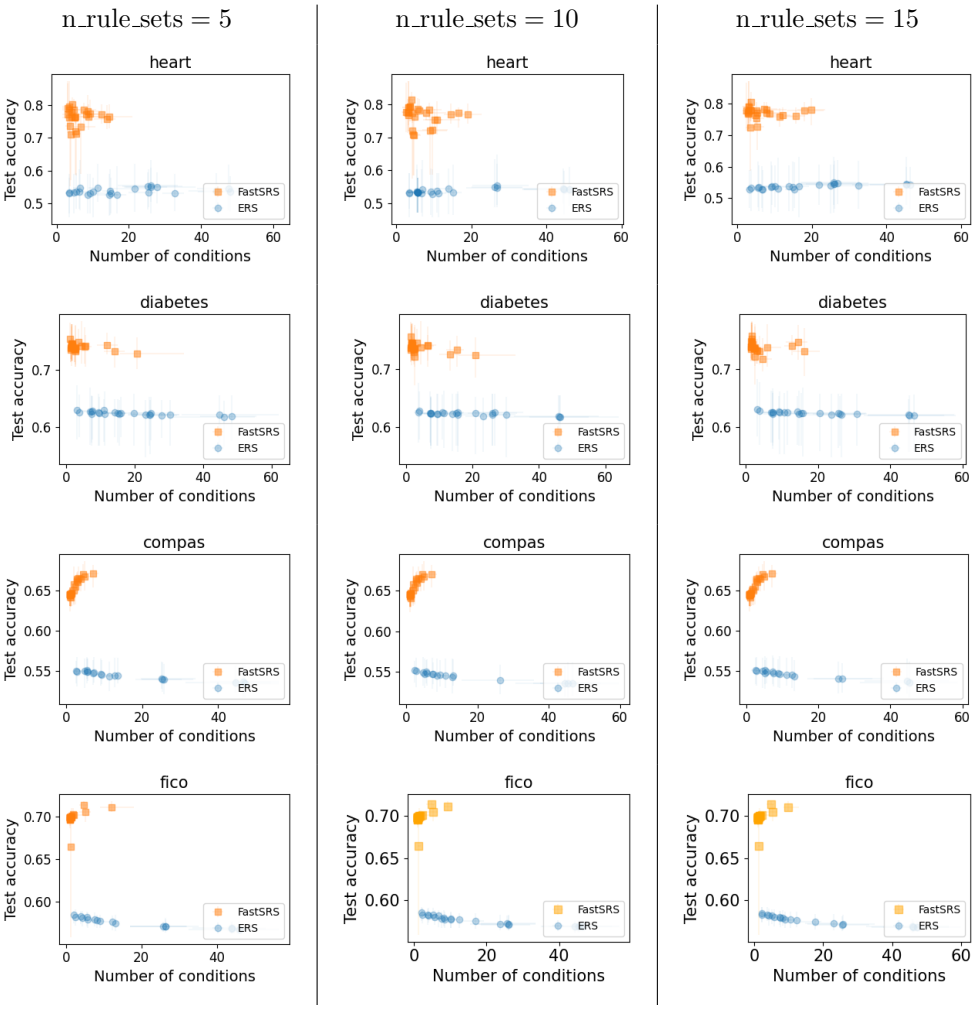

Fig. J13 Average accuracy test versus average number of conditions across the rule sets in the Rashomon set for all configuration tested (1/2).

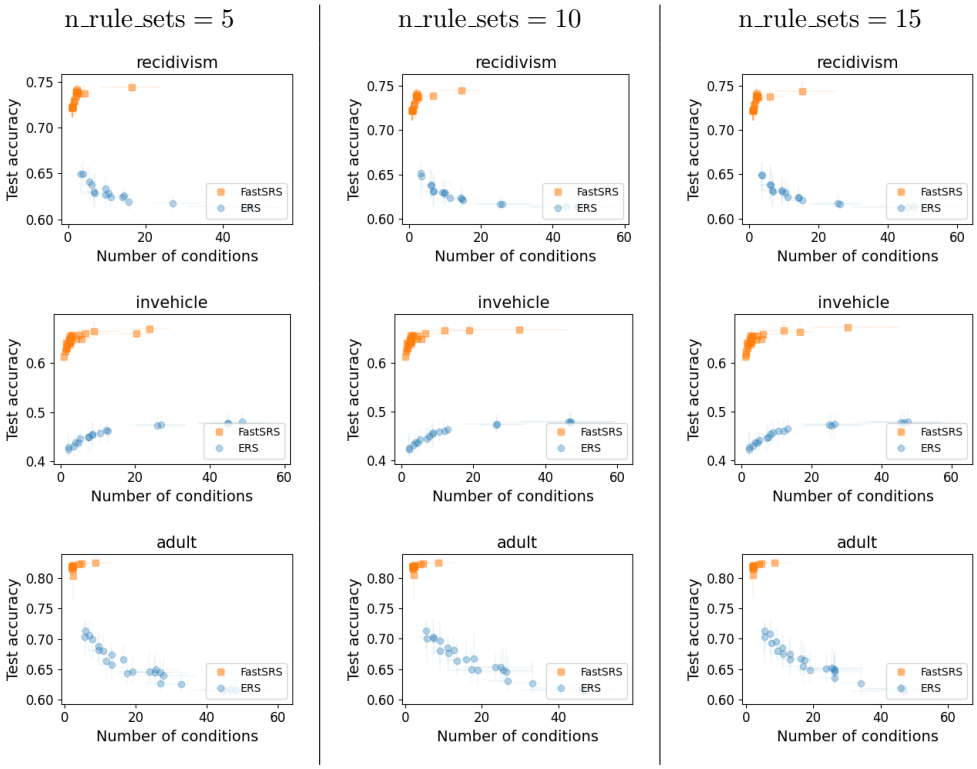

Fig. J14 Average accuracy test versus average number of conditions across the rule sets in the Rashomon set for all configuration tested (2/2).

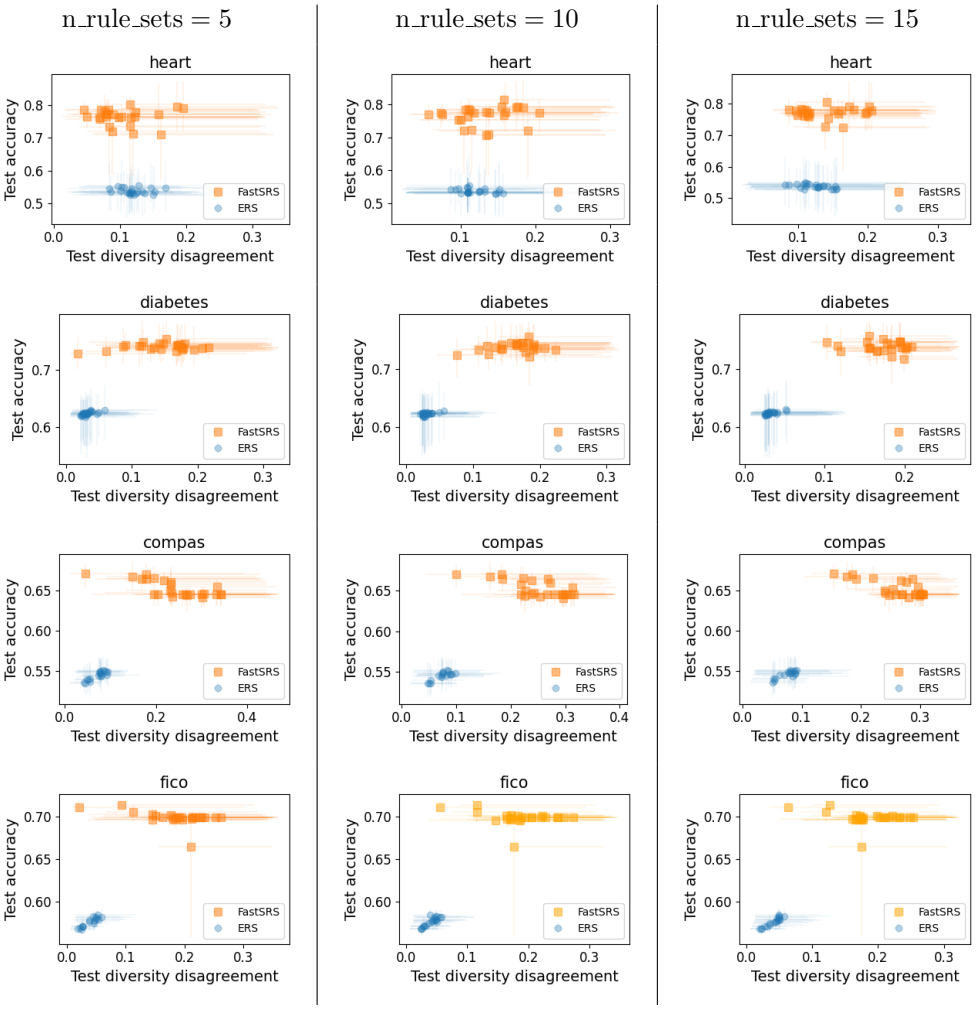

Fig. J15 Average accuracy test versus diversity disagreement (test) across the rule sets in the Rashomon set for all configuration tested (1/2).

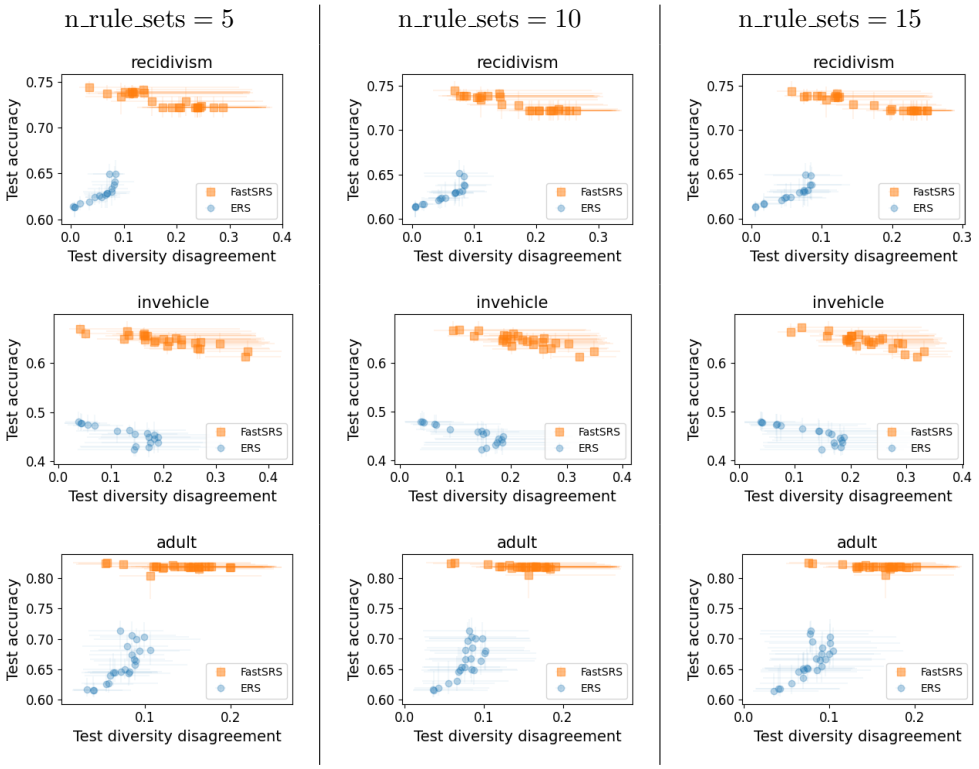

Fig. J16 Average accuracy test versus diversity disagreement (test) across the rule sets in the Rashomon set for all configurations tested (2/2).

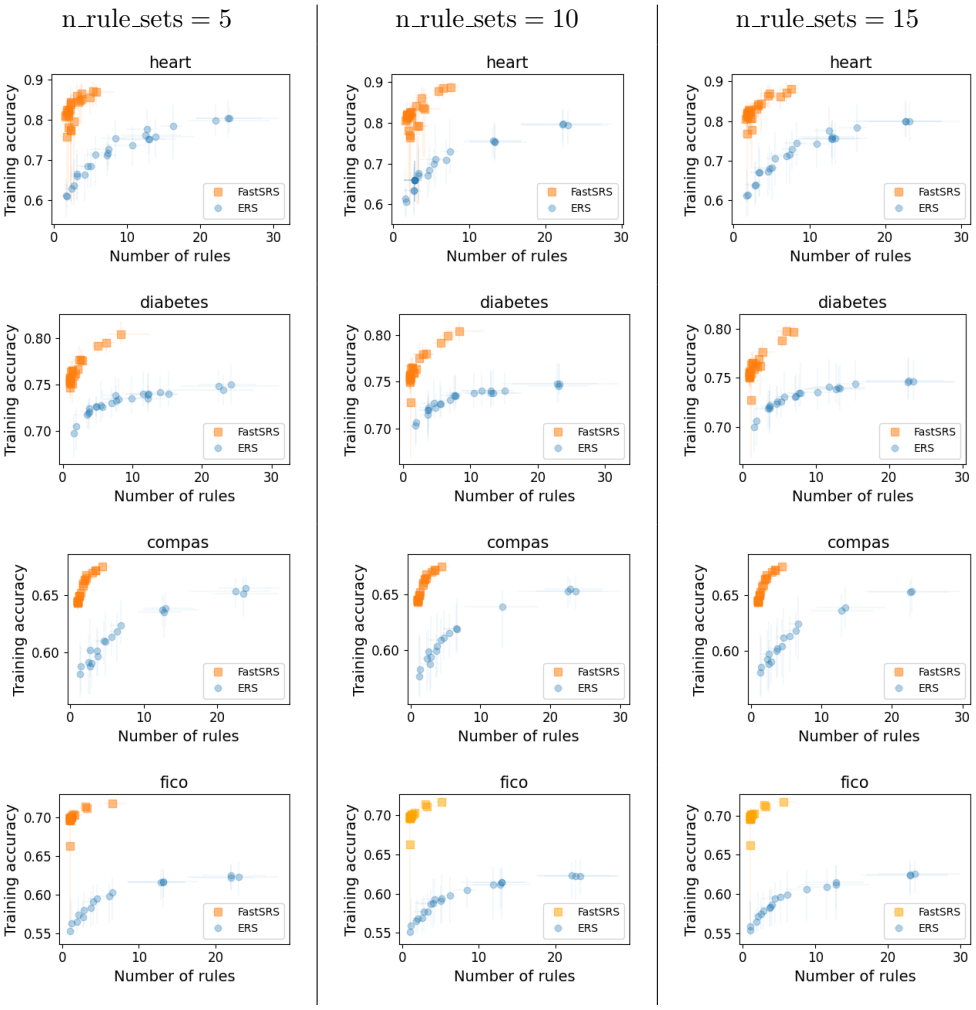

Fig. J17 Average training accuracy versus average number of rules across the rule sets in the Rashomon set for all configurations tested (1/2).

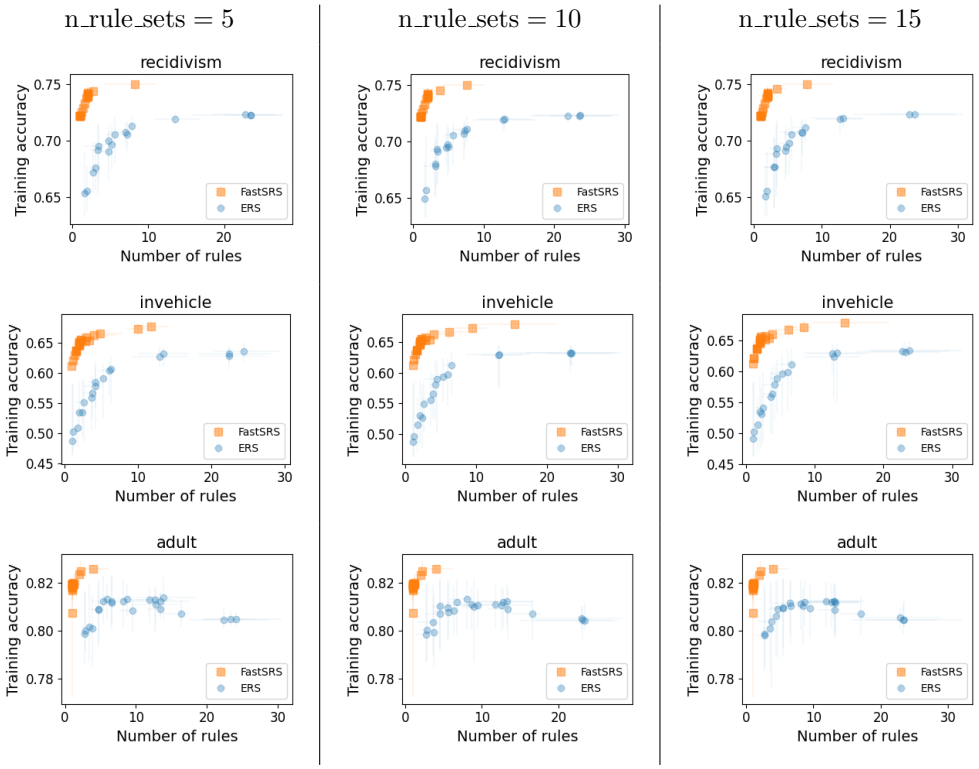

Fig. J18 Average training accuracy versus average number of rules across the rule sets in the Rashomon set for all configurations tested (2/2).

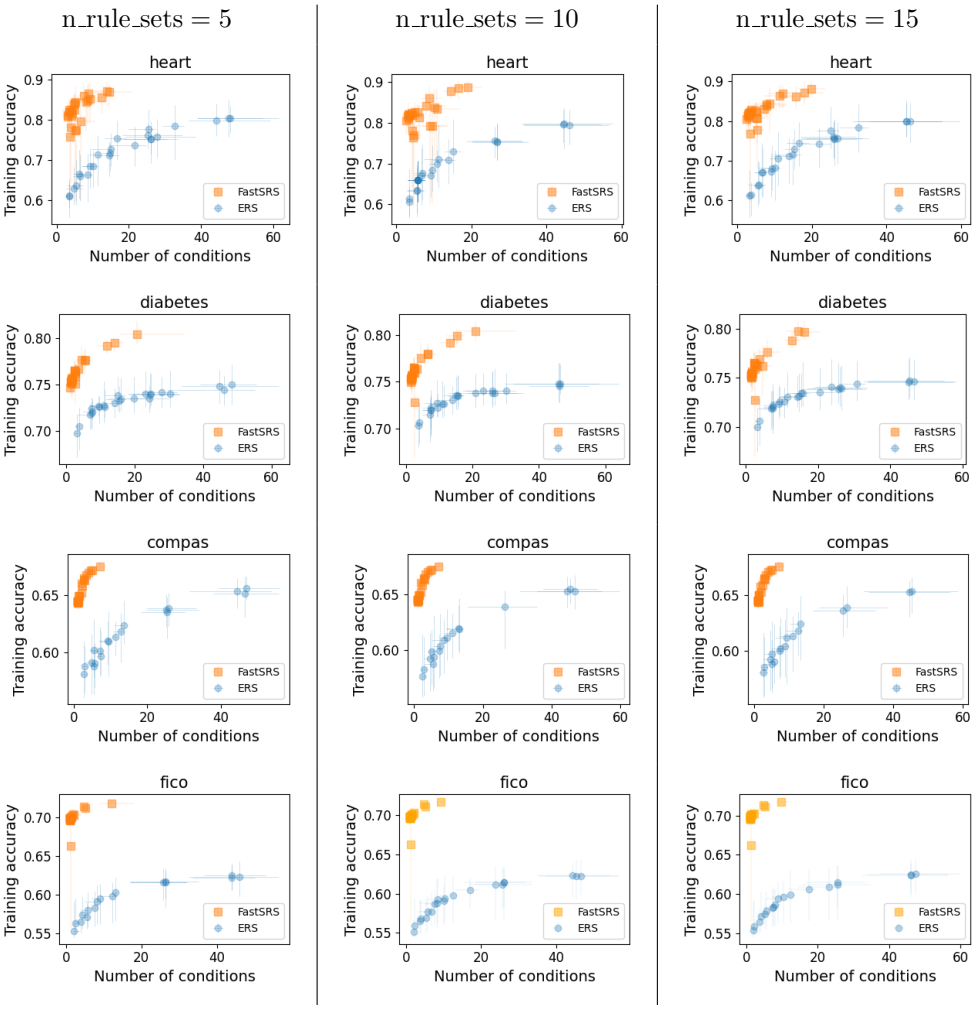

Fig. J19 Average training accuracy versus average number of conditions across the rule sets in the Rashomon set for all configurations tested (1/2).

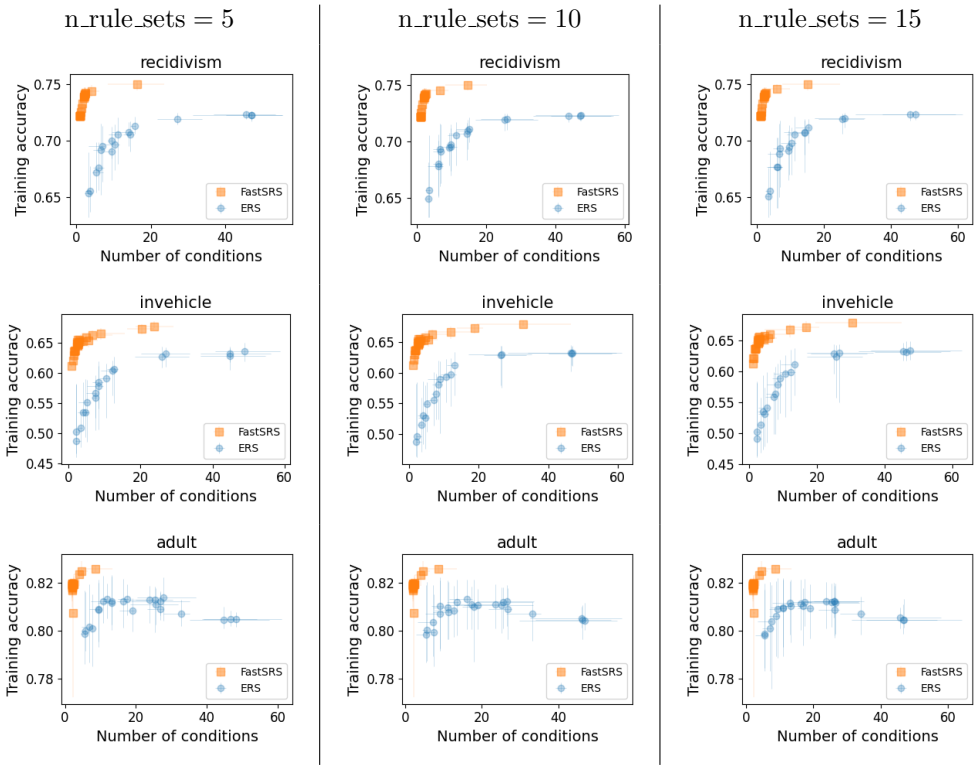

Fig. J20 Average training accuracy versus average number of conditions across the rule sets in the Rashomon set for all configurations tested (2/2).

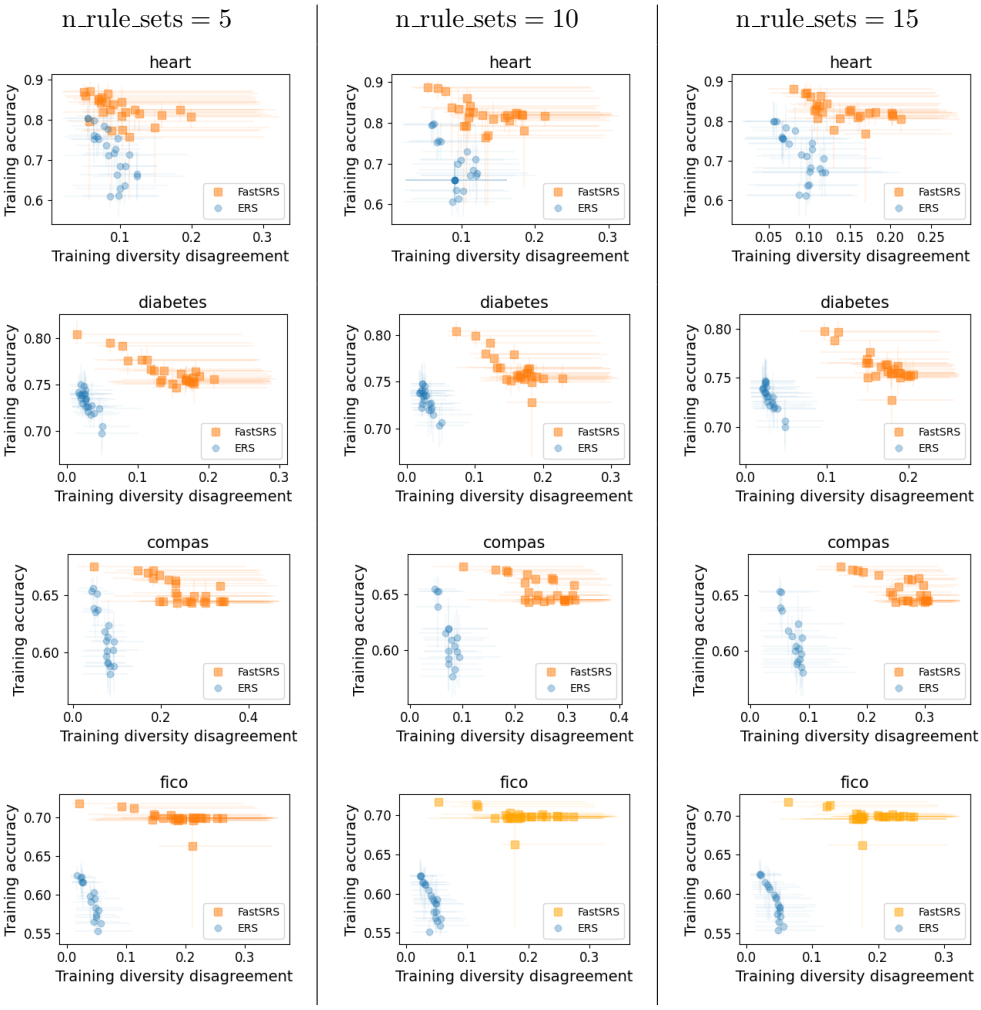

Fig. J21 Average training accuracy versus diversity disagreement (train) across the rule sets in the Rashomon set for all configurations tested (1/2).

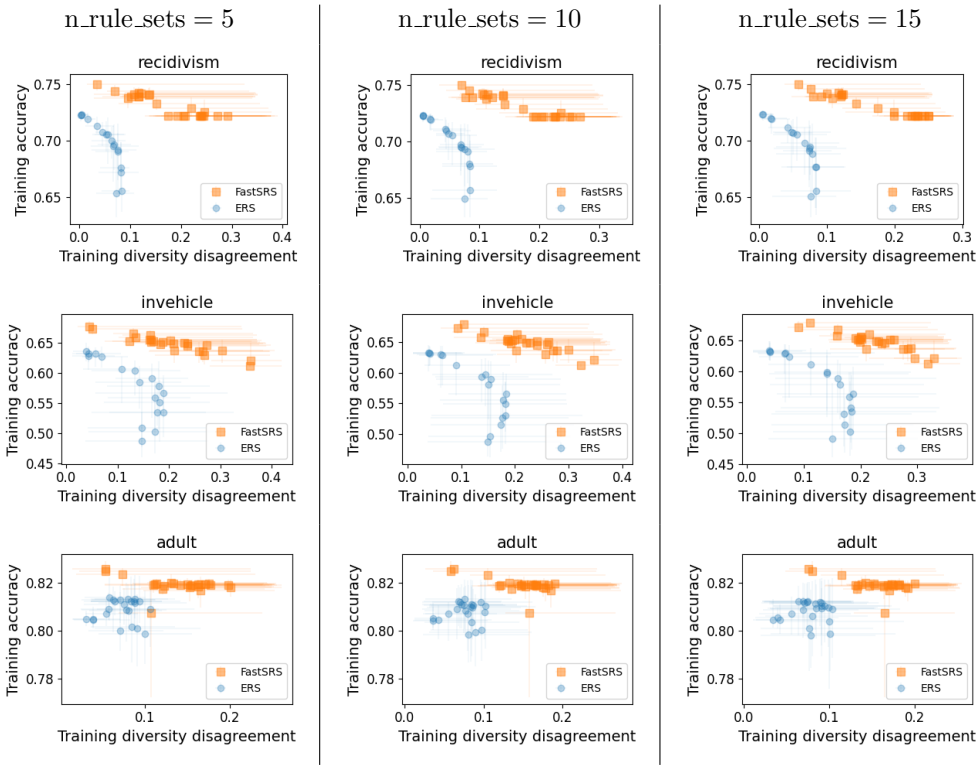

Fig. J22 Average training accuracy versus diversity disagreement (train) across the rule sets in the Rashomon set for all configurations tested (2/2).

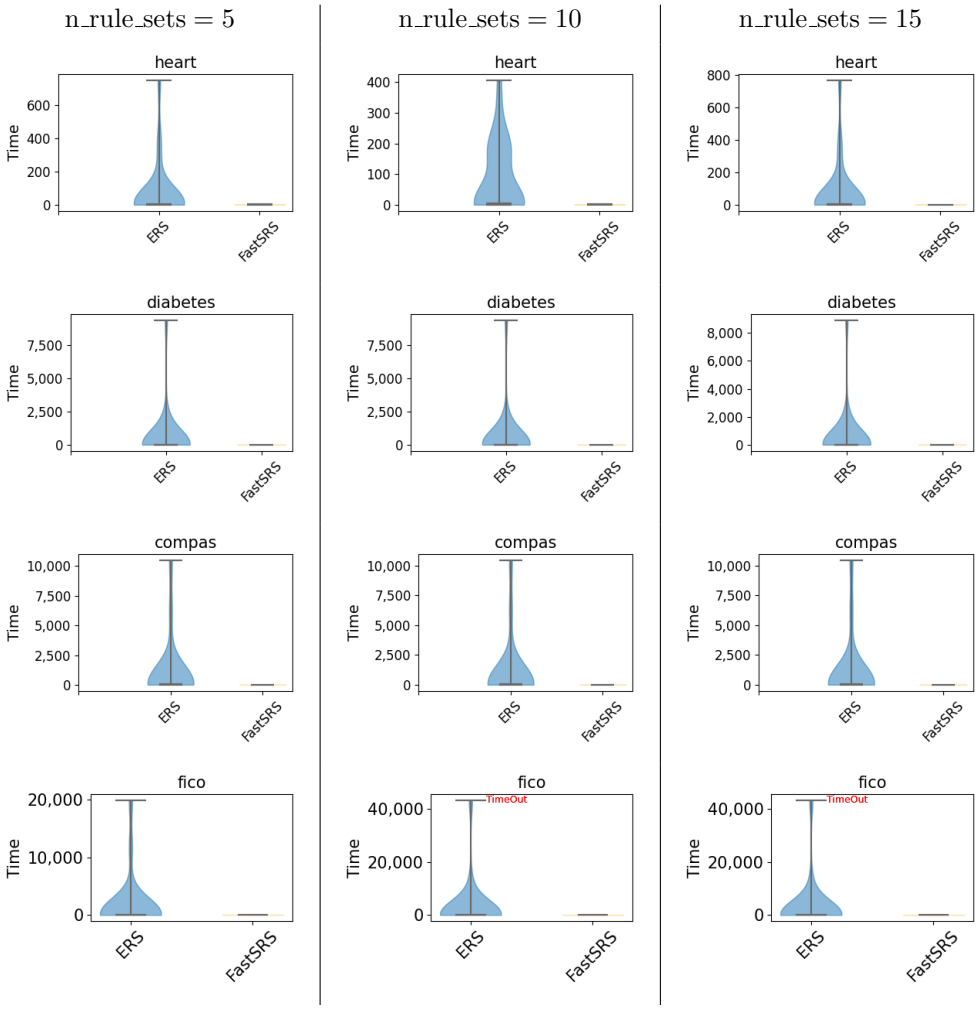

Fig. J23 Computation time across the rule sets in the Rashomon set for all configurations tested (1/2).

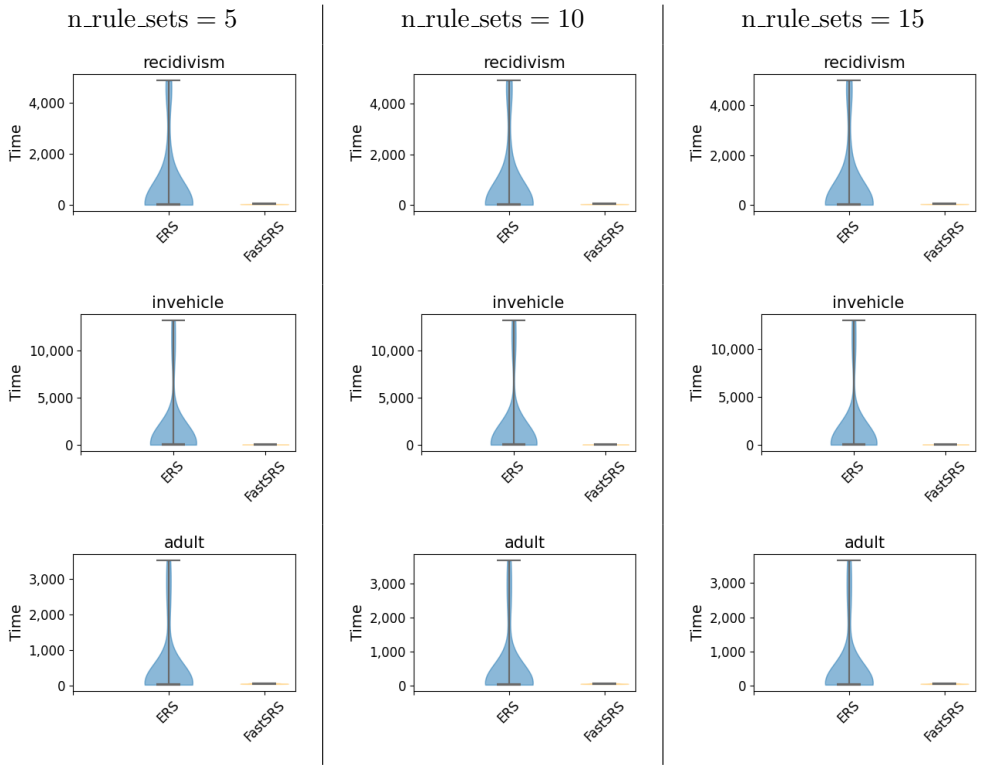

Fig. J24 Computation time across the rule sets in the Rashomon set for all configurations tested (2/2).

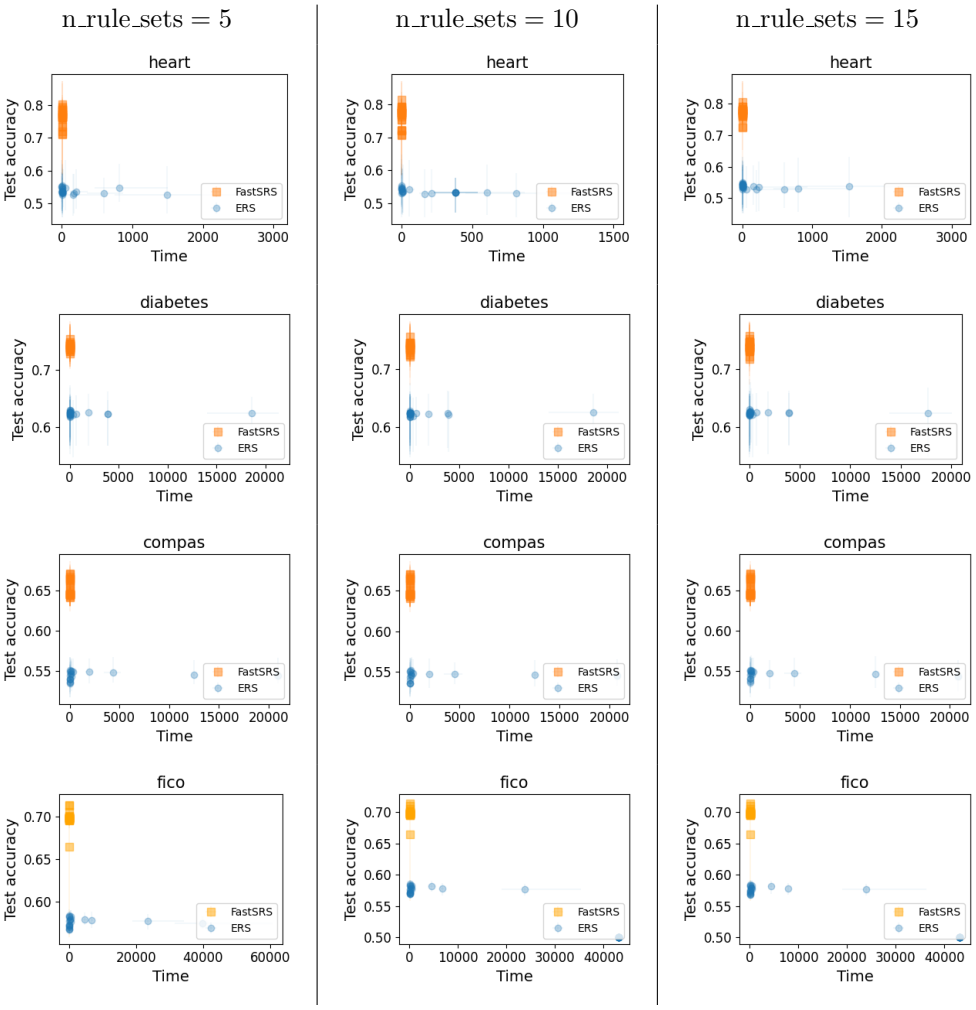

Fig. J25 Average test accuracy versus computation time across the rule sets in the Rashomon set for all configurations tested (1/2).

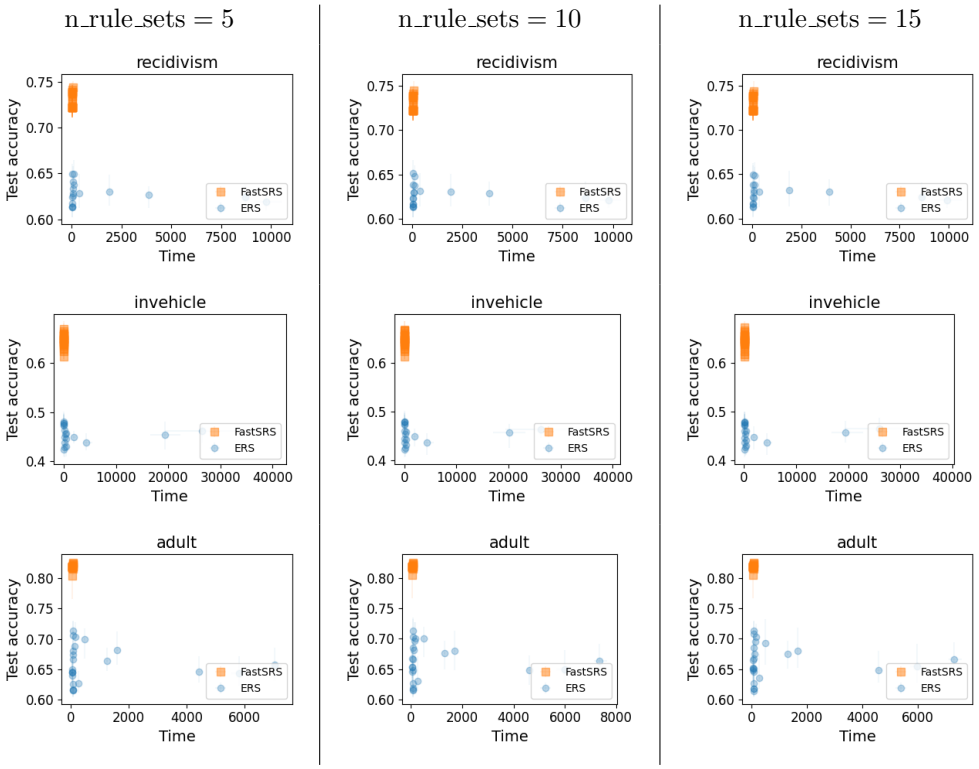

Fig. J26 Average test accuracy versus computation time across the rule sets in the Rashomon set for all configurations tested (2/2).

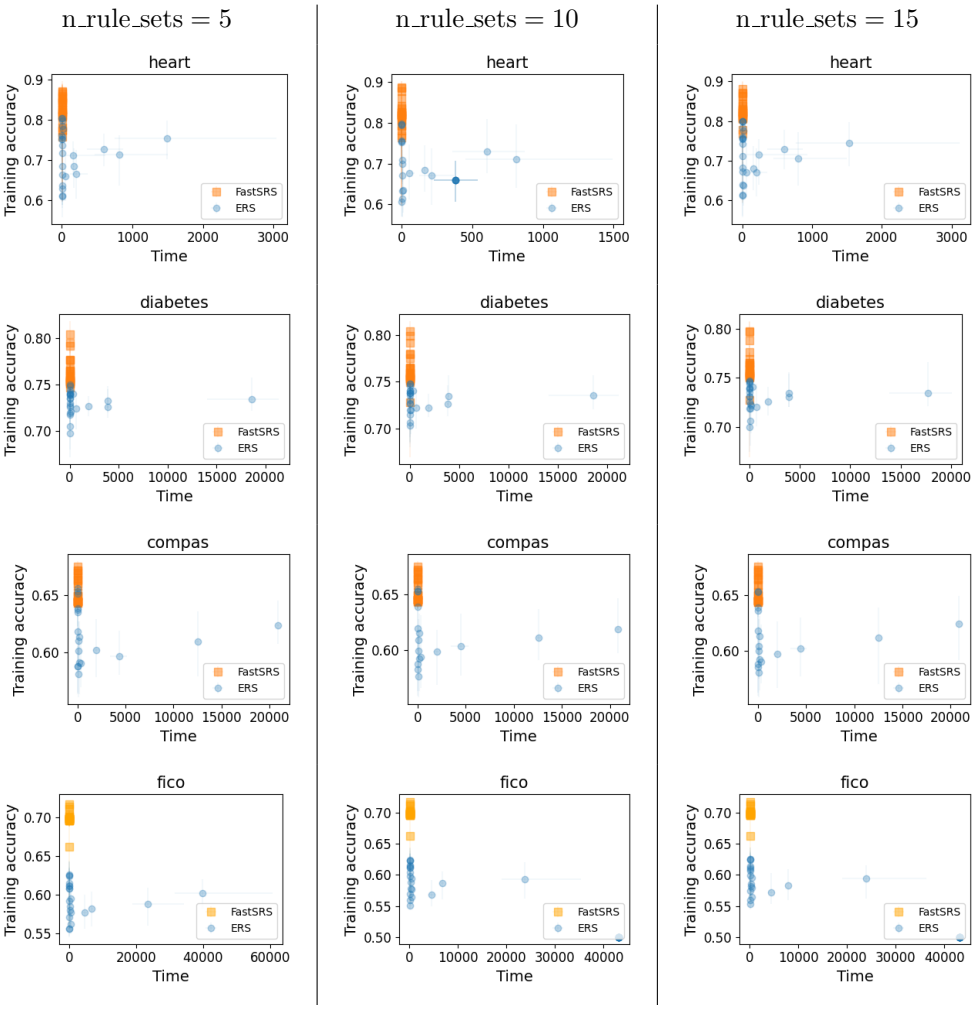

Fig. J27 Average training accuracy versus computation time across the rule sets in the Rashomon set for all configurations tested (1/2).

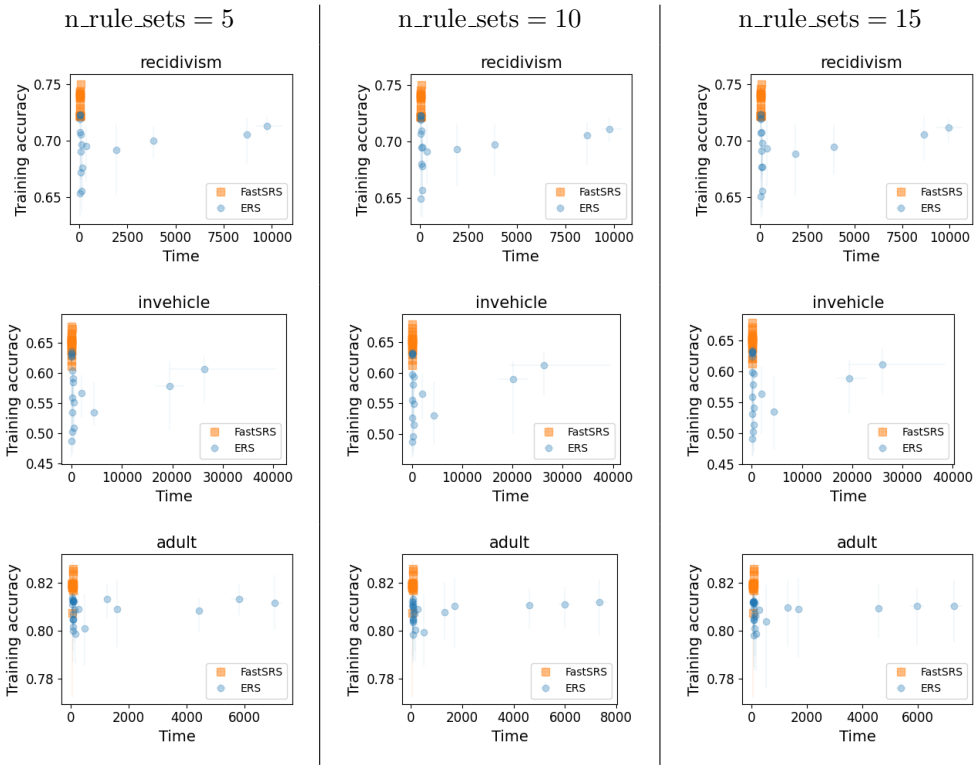

Fig. J28 Average training accuracy versus computation time across the rule sets in the Rashomon set for all configurations tested (2/2).

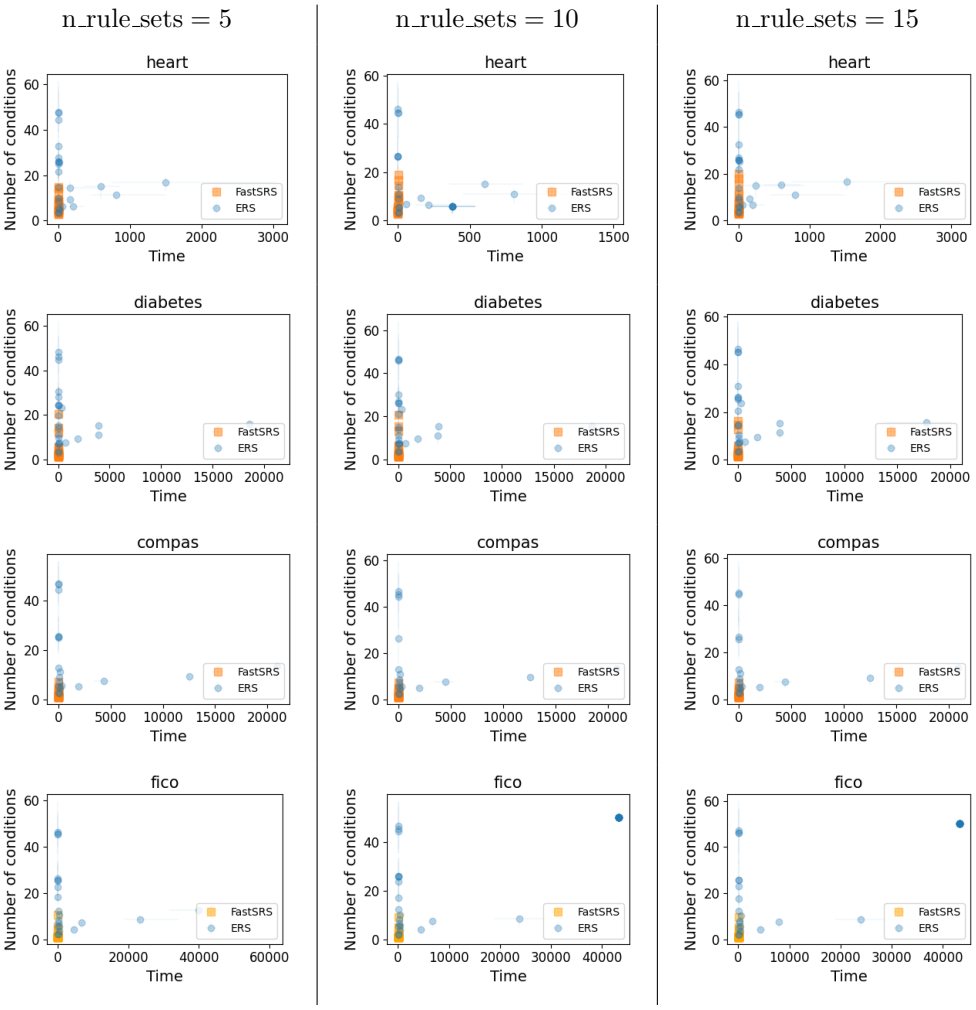

Fig. J29 Number of conditions across the rule sets in the Rashomon set vs computation time for all configurations tested (1/2).

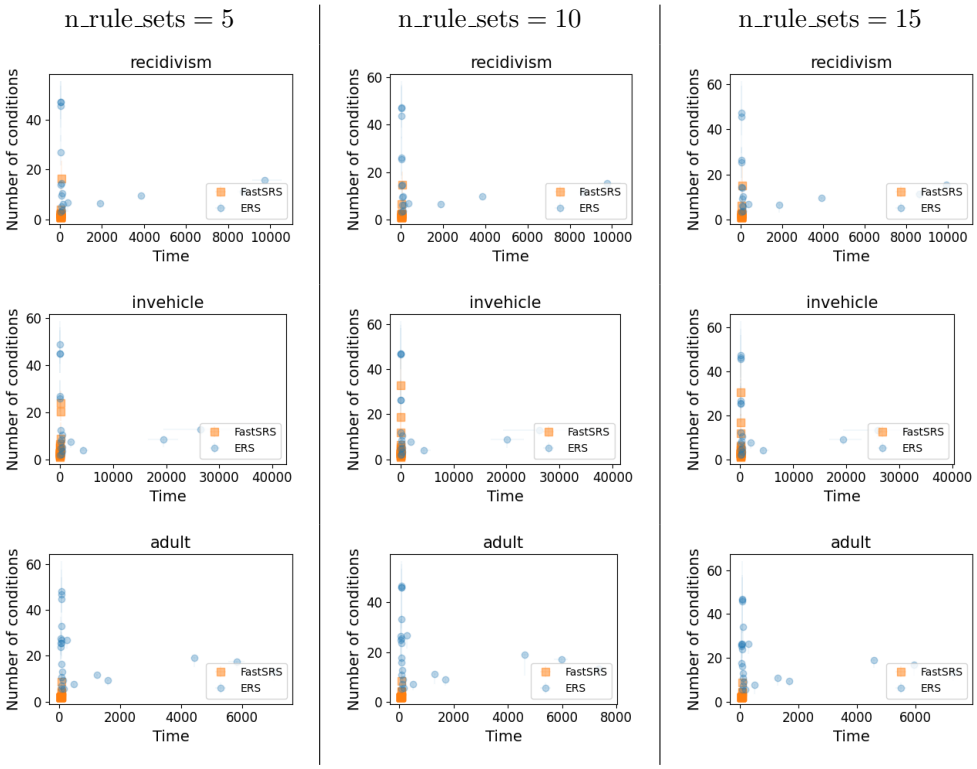

Fig. J30 Number of conditions across the rule sets in the Rashomon set vs computation time for all configurations tested (2/2).

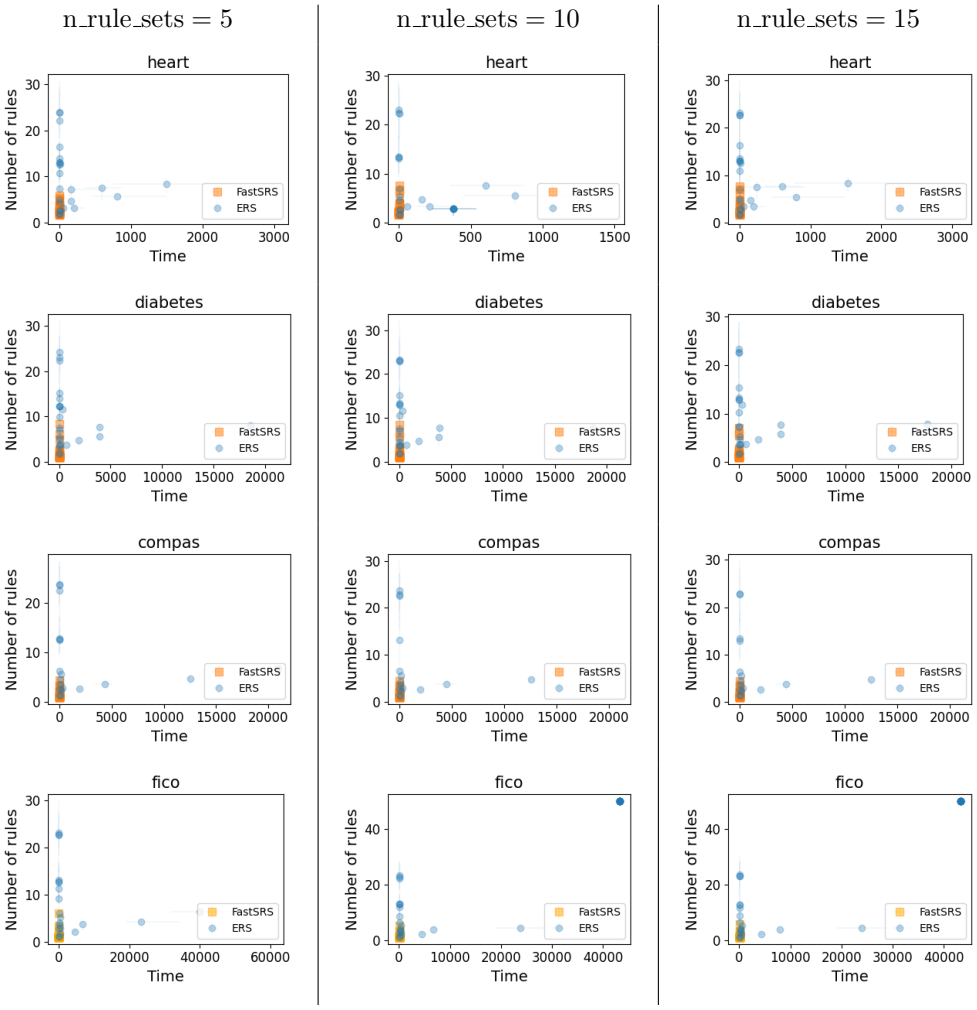

Fig. J31 Number of rules across the rule sets in the Rashomon set vs computation time for all configurations tested (1/2).

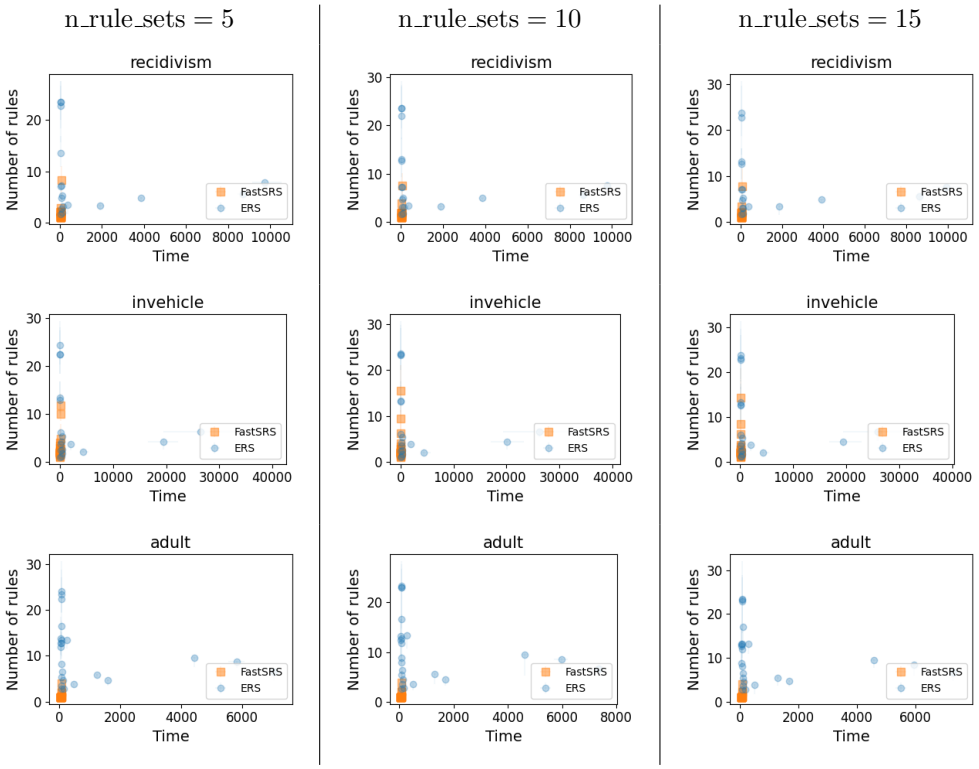

Fig. J32 Number of rules across the rule sets in the Rashomon set vs computation time for all configurations tested (2/2).

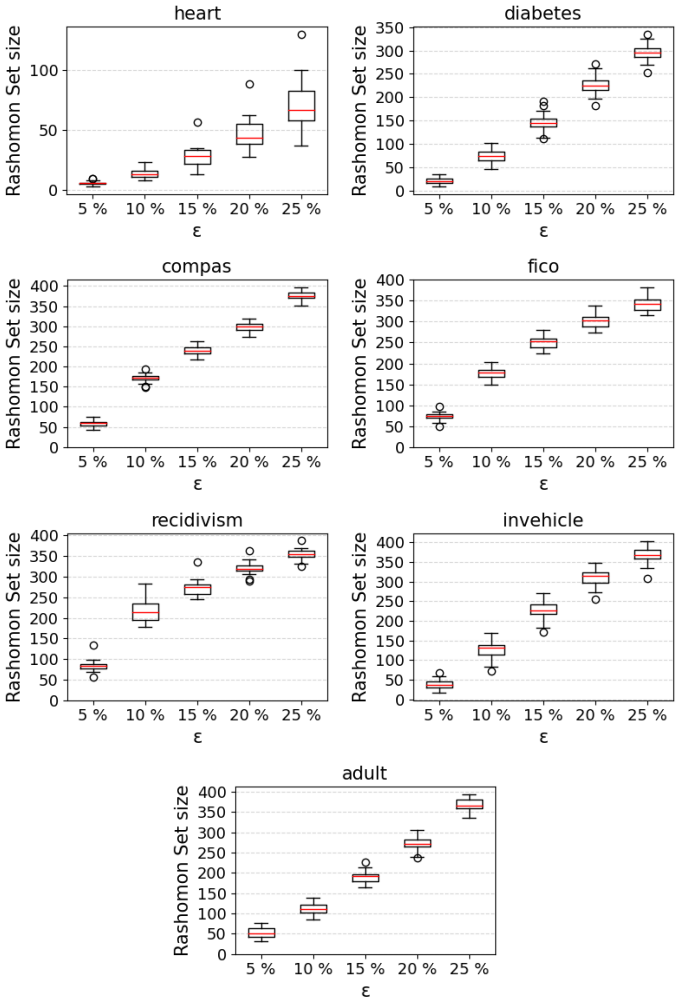

Fig. J33 Rashomon set sizes for FastSRS with $N_{\text{iter}} = 500$ as a function of ϵ.

J.1 Results on structural diversity

Figures J34 through J61 show our results on structural diversity.

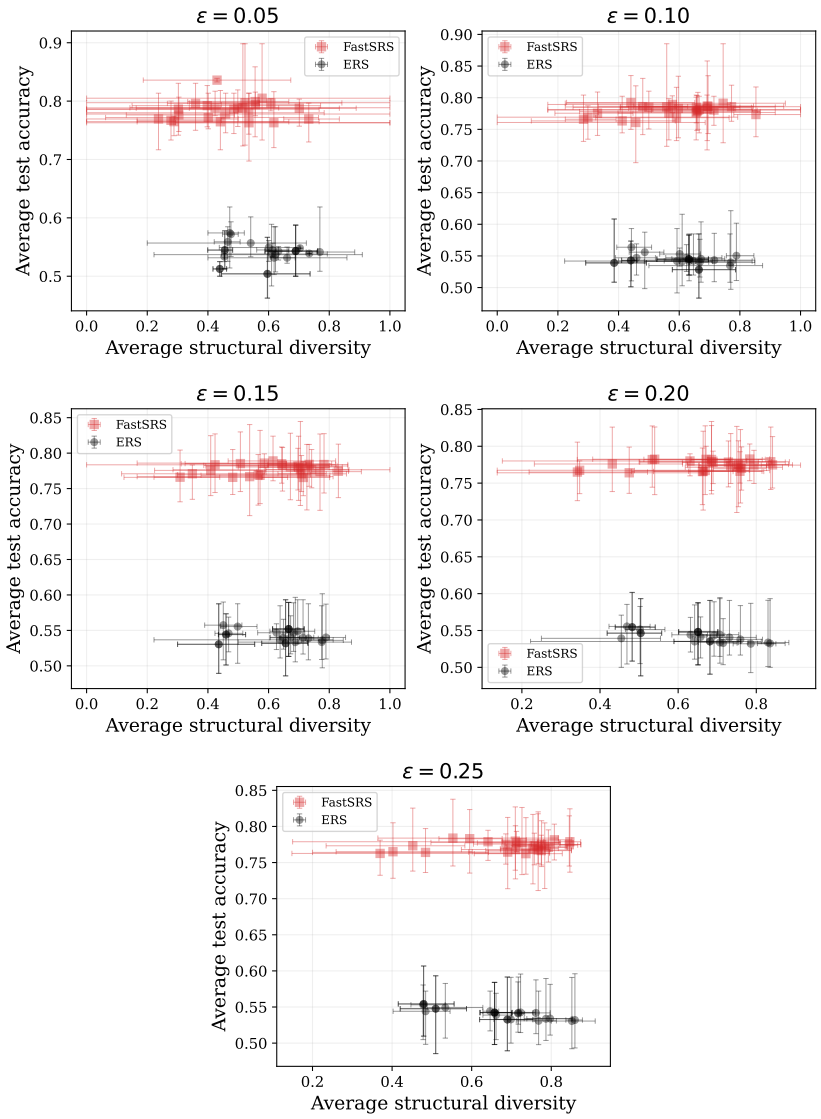

Fig. J34 Comparison between FastSRS and ERS in terms of test accuracy and average structural diversity for heart dataset.

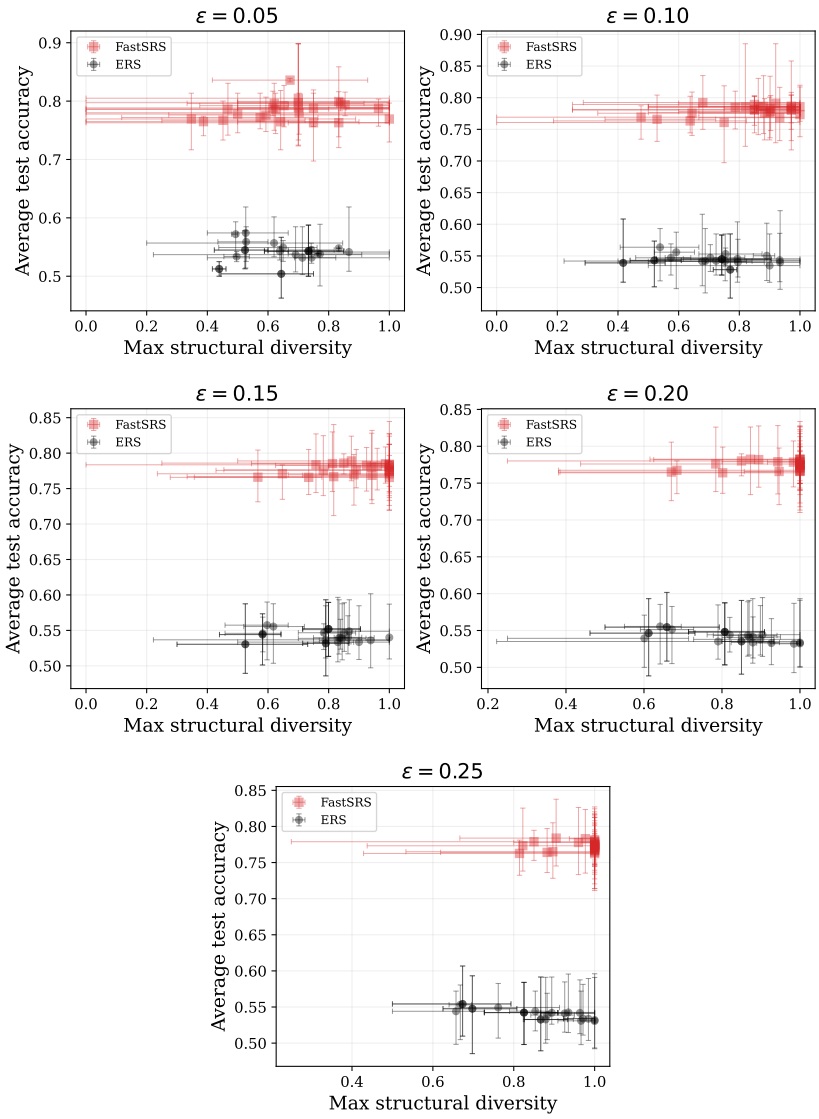

Fig. J35 Comparison between FastSRS and ERS in terms of test accuracy and maximum structural diversity for heart dataset.

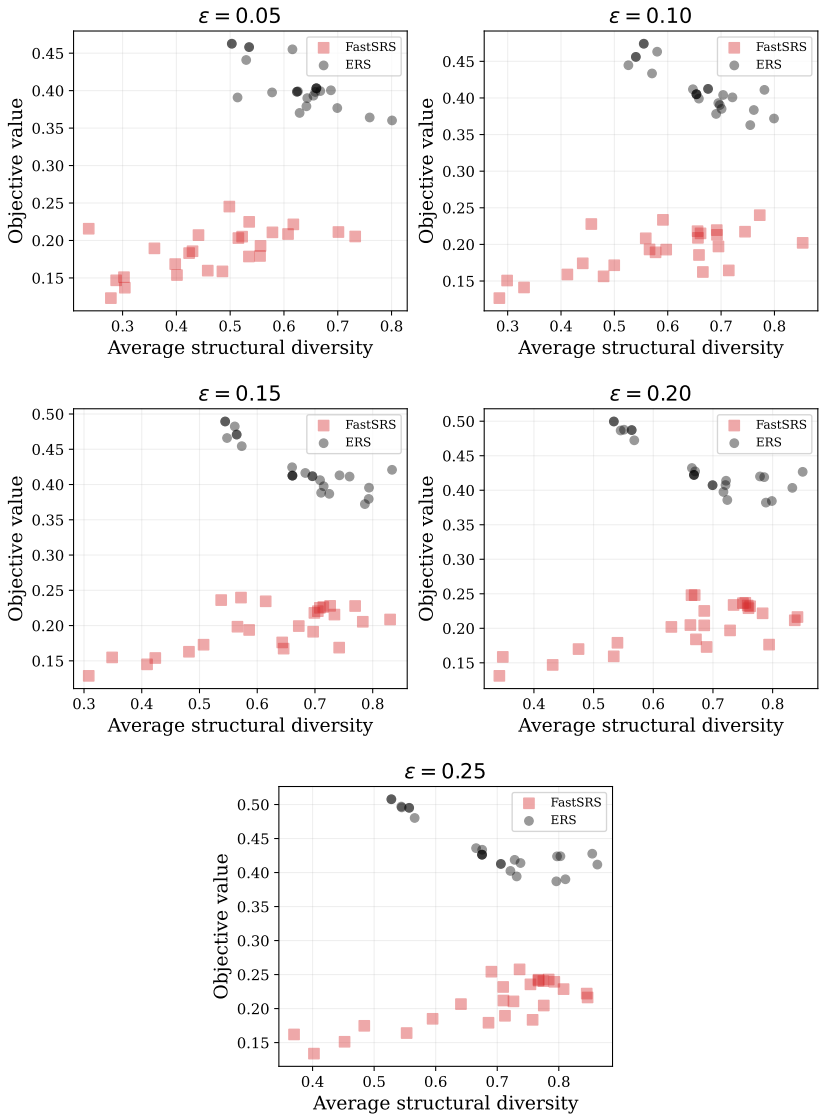

Fig. J36 Comparison between FastSRS and ERS in terms of objective value and average structural diversity for heart dataset.

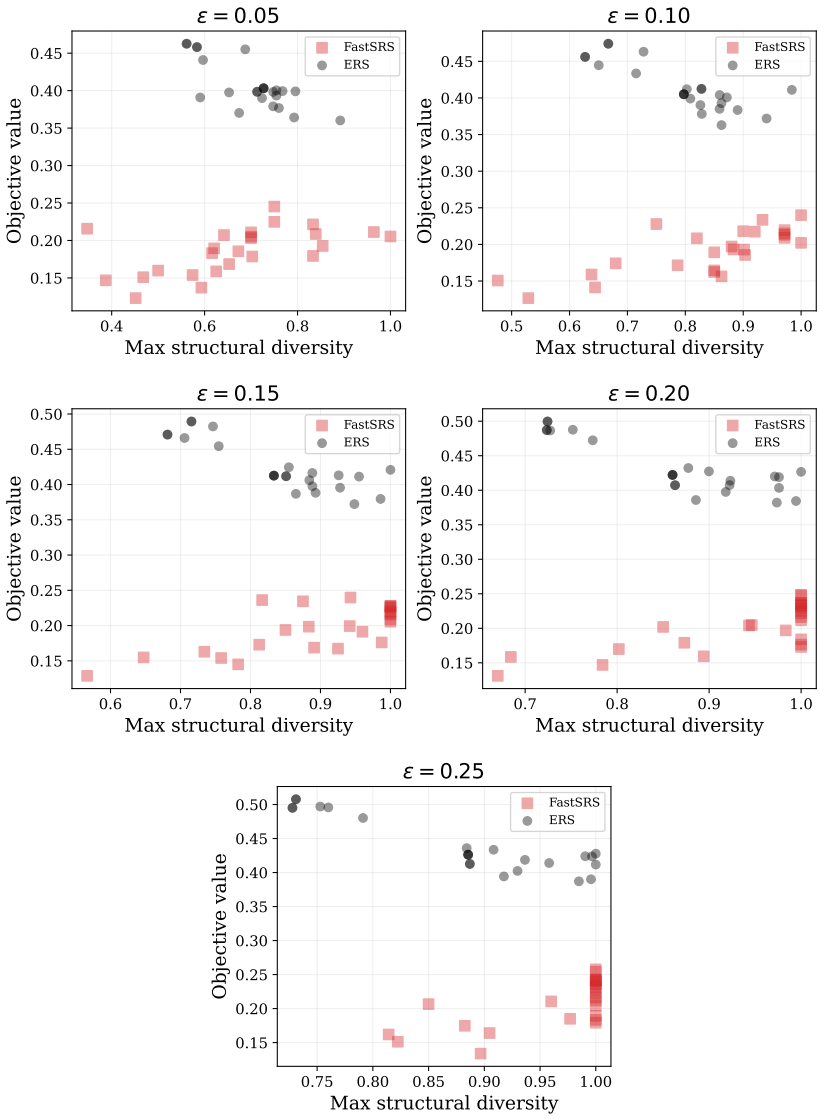

Fig. J37 Comparison between FastSRS and ERS in terms of objective value and maximum structural diversity for heart dataset.

diabetes

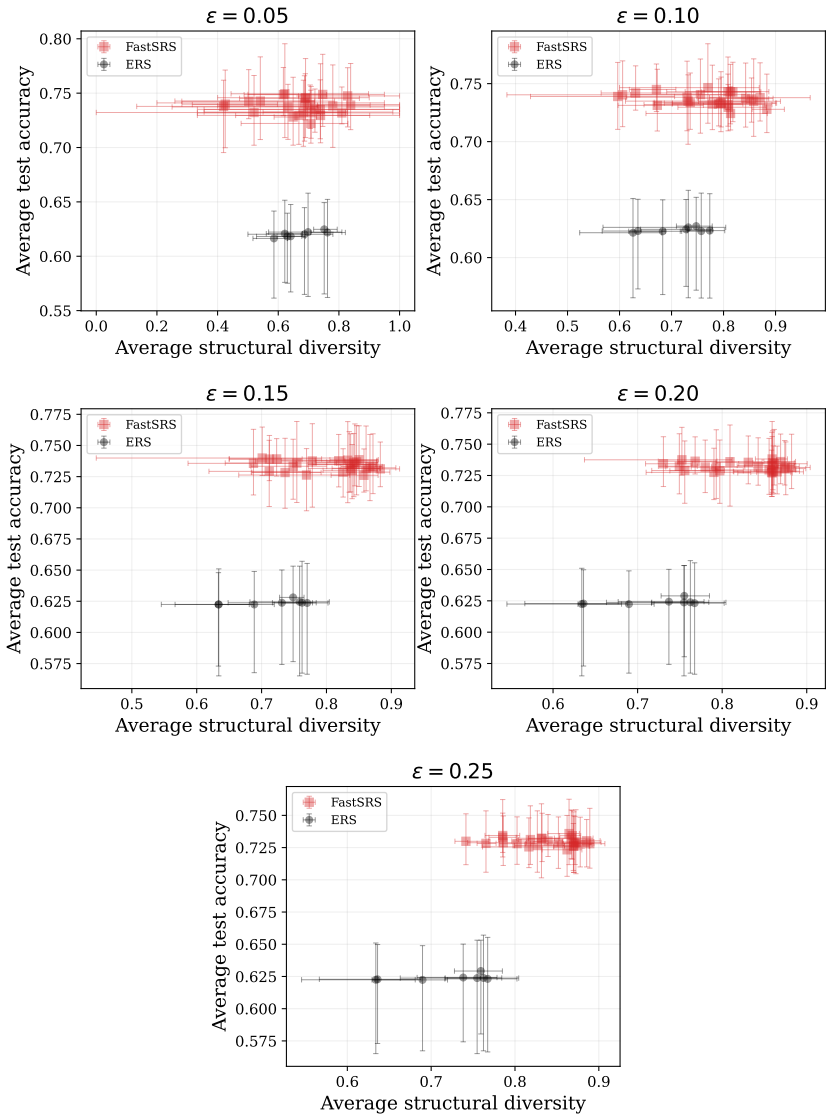

Fig. J38 Comparison between FastSRS and ERS in terms of test accuracy and average structural diversity for diabetes dataset.

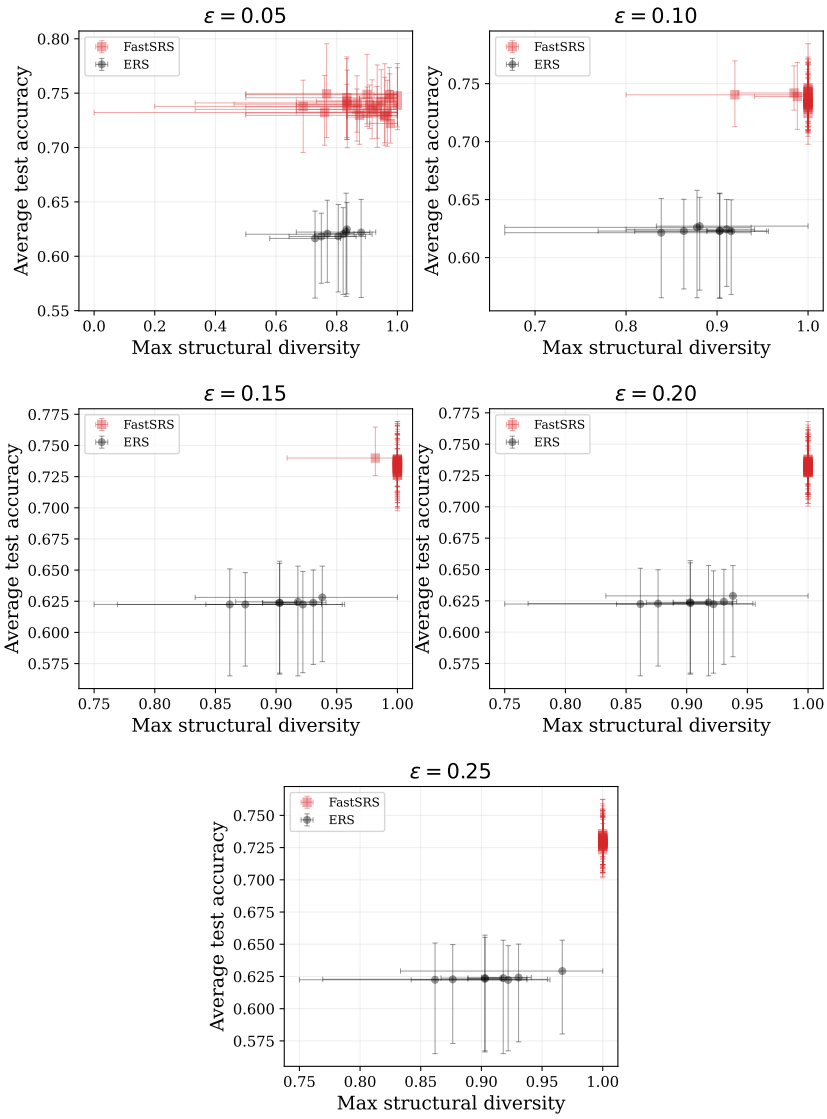

Fig. J39 Comparison between FastSRS and ERS in terms of test accuracy and maximum structural diversity for diabetes dataset.

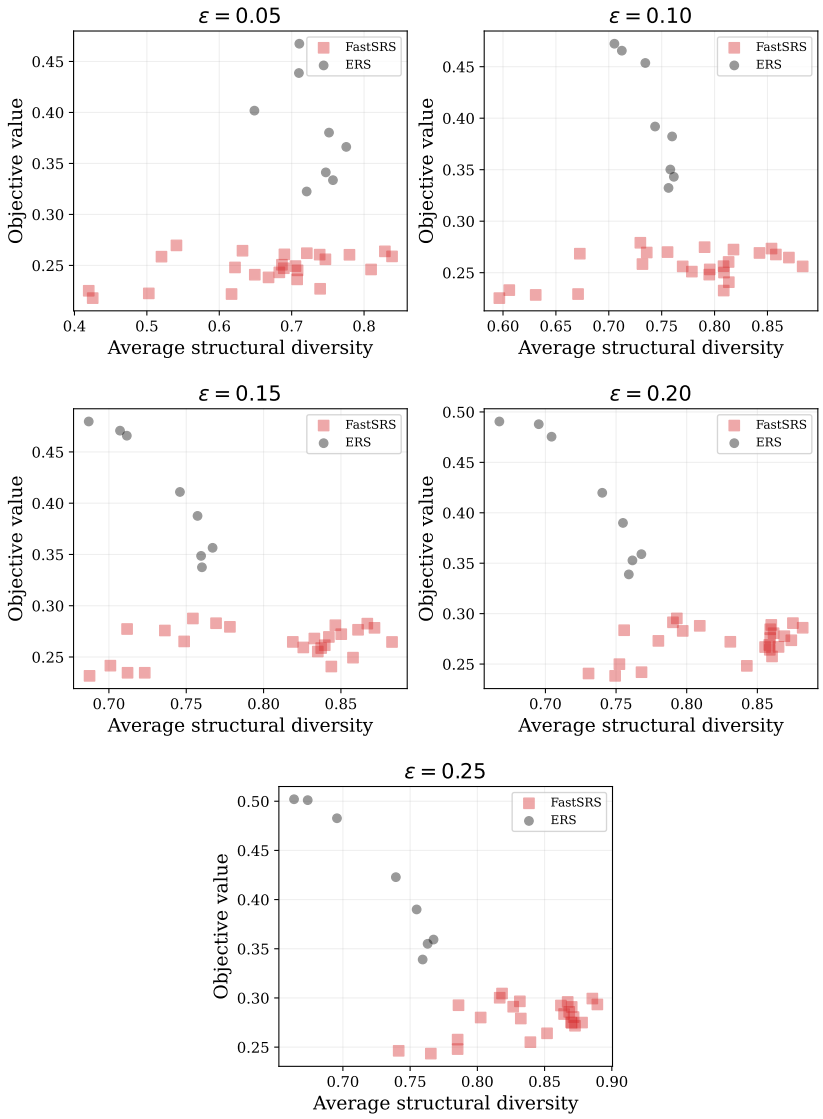

Fig. J40 Comparison between FastSRS and ERS in terms of objective value and average structural diversity for diabetes dataset.

diabetes

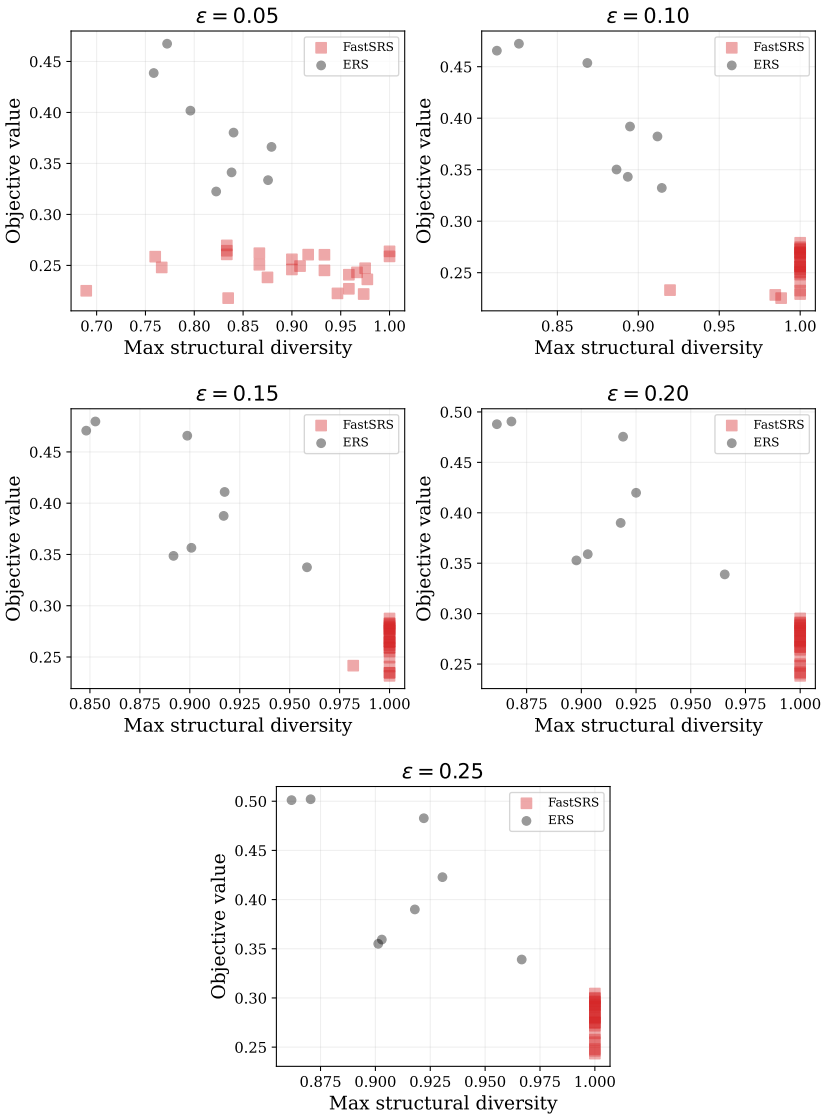

Fig. J41 Comparison between FastSRS and ERS in terms of objective value and maximum structural diversity for diabetes dataset.

compas

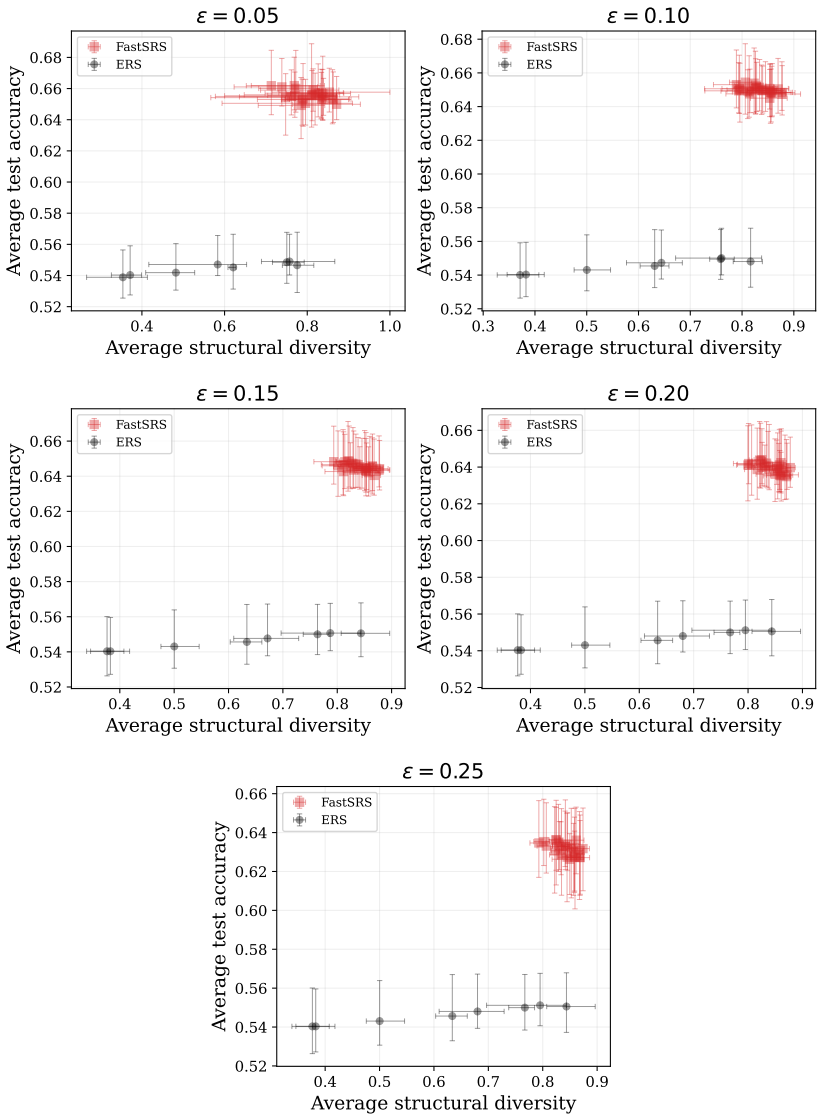

Fig. J42 Comparison between FastSRS and ERS in terms of test accuracy and average structural diversity for compas dataset.

compas

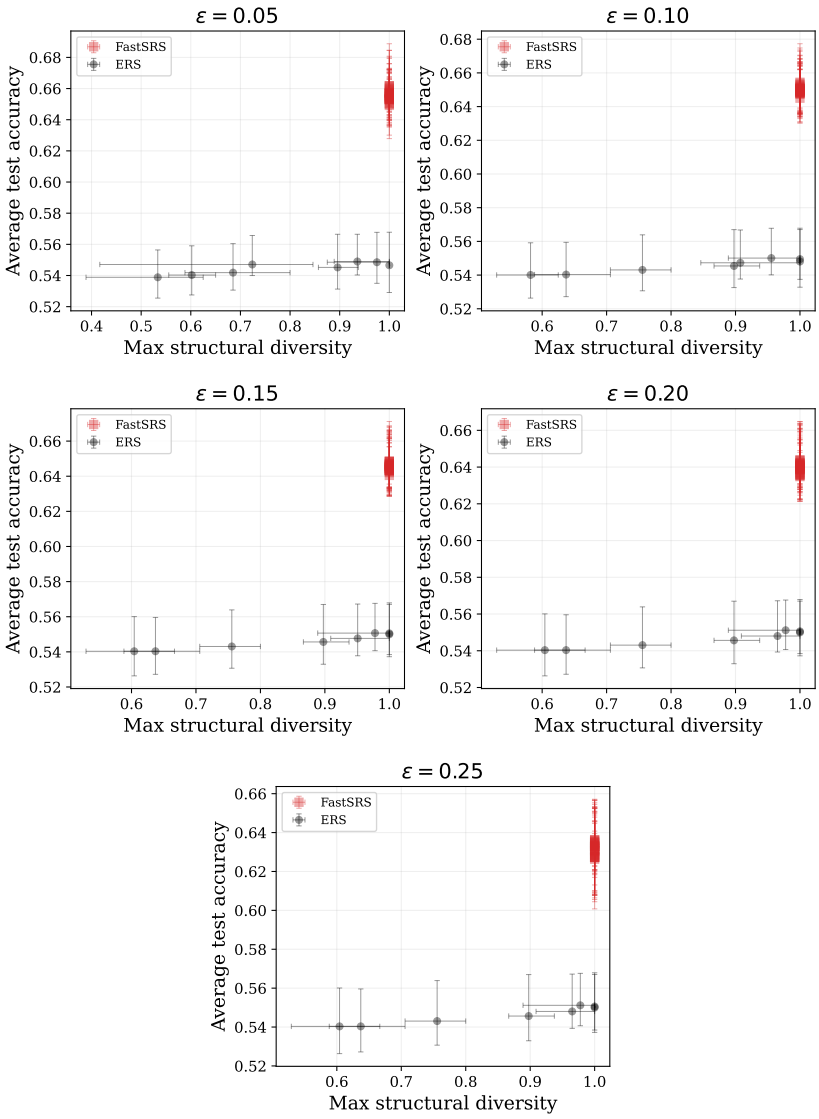

Fig. J43 Comparison between FastSRS and ERS in terms of test accuracy and maximum structural diversity for compas dataset.

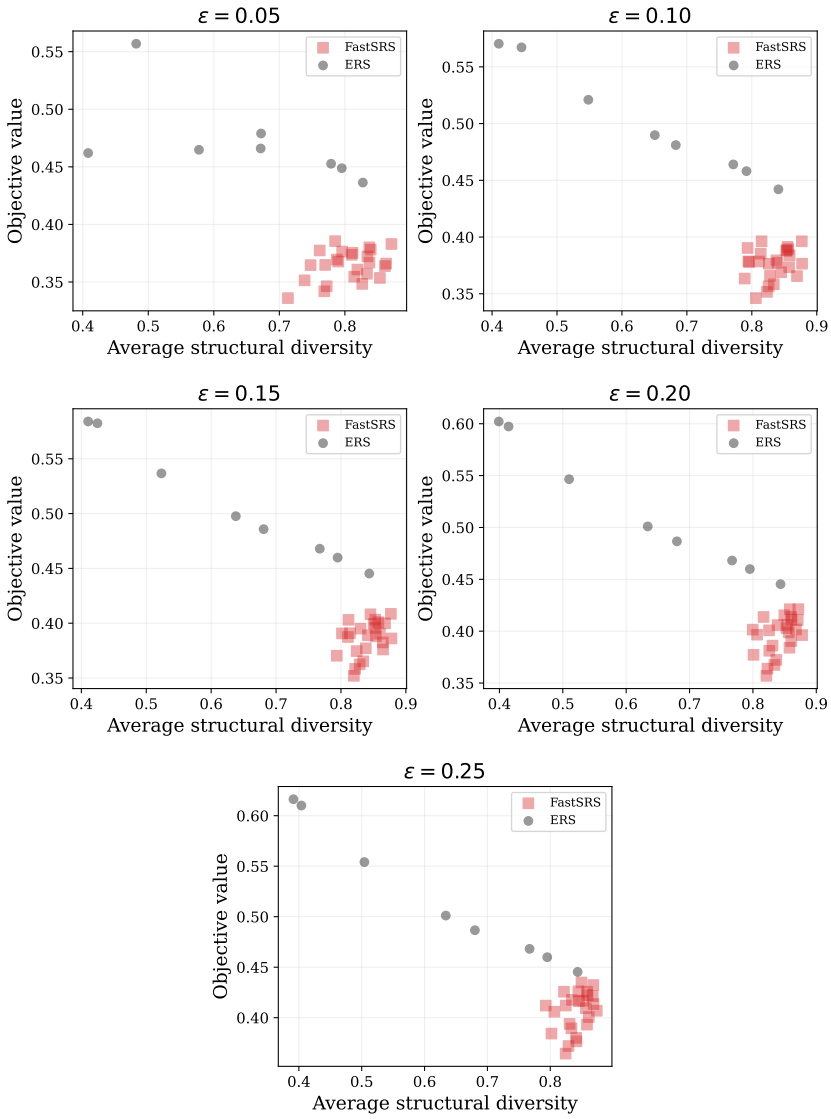

Fig. J44 Comparison between FastSRS and ERS in terms of objective value and average structural diversity for compas dataset.

compas

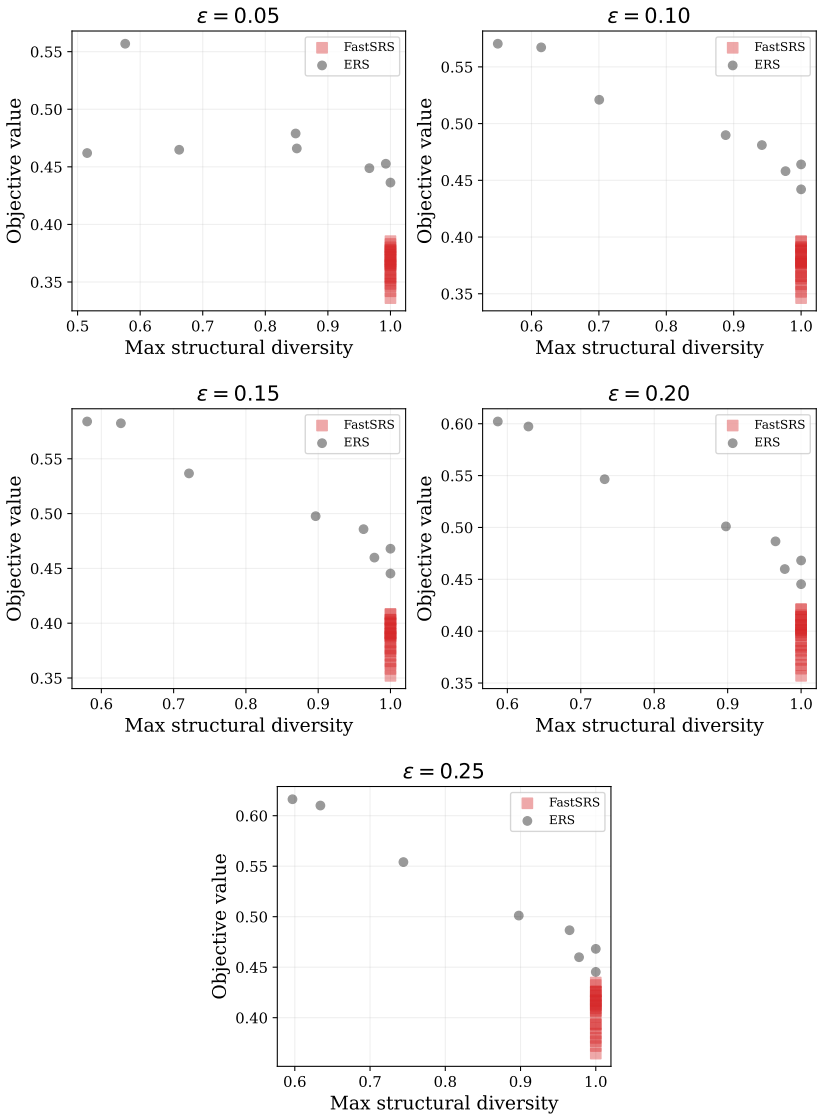

Fig. J45 Comparison between FastSRS and ERS in terms of objective value and maximum structural diversity for compas dataset.

fico

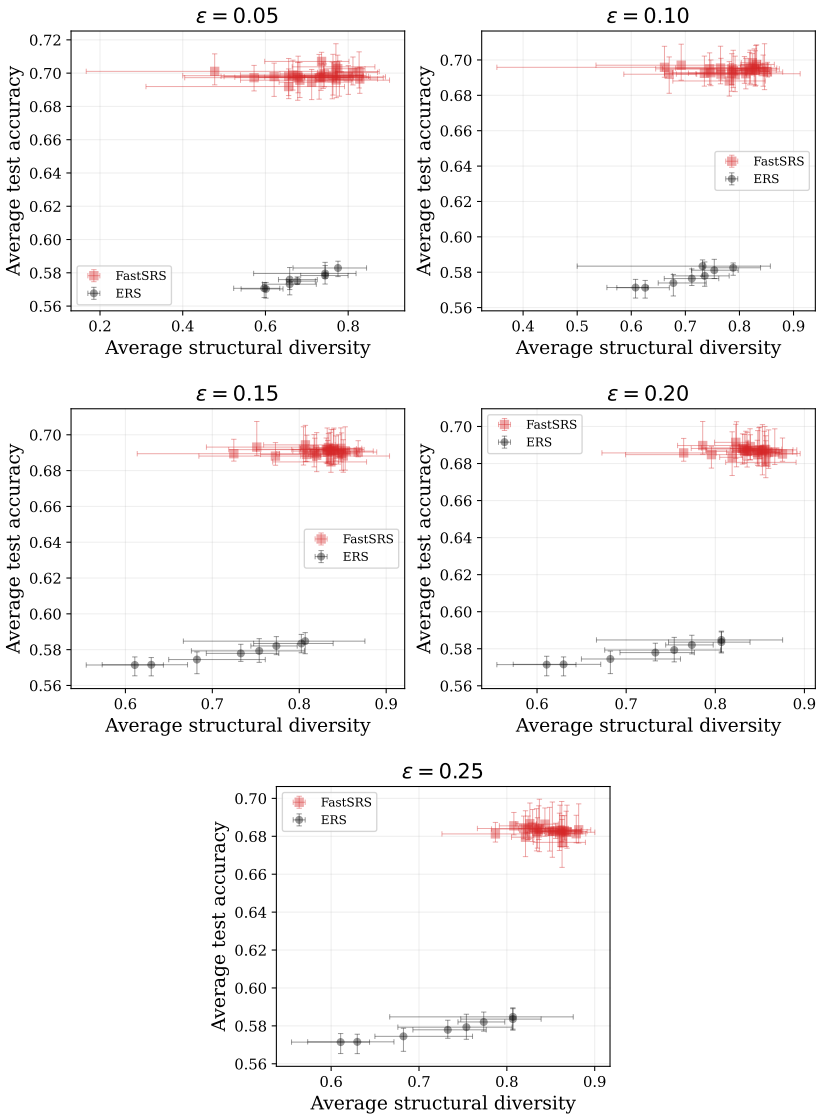

Fig. J46 Comparison between FastSRS and ERS in terms of test accuracy and average structural diversity for fico dataset.

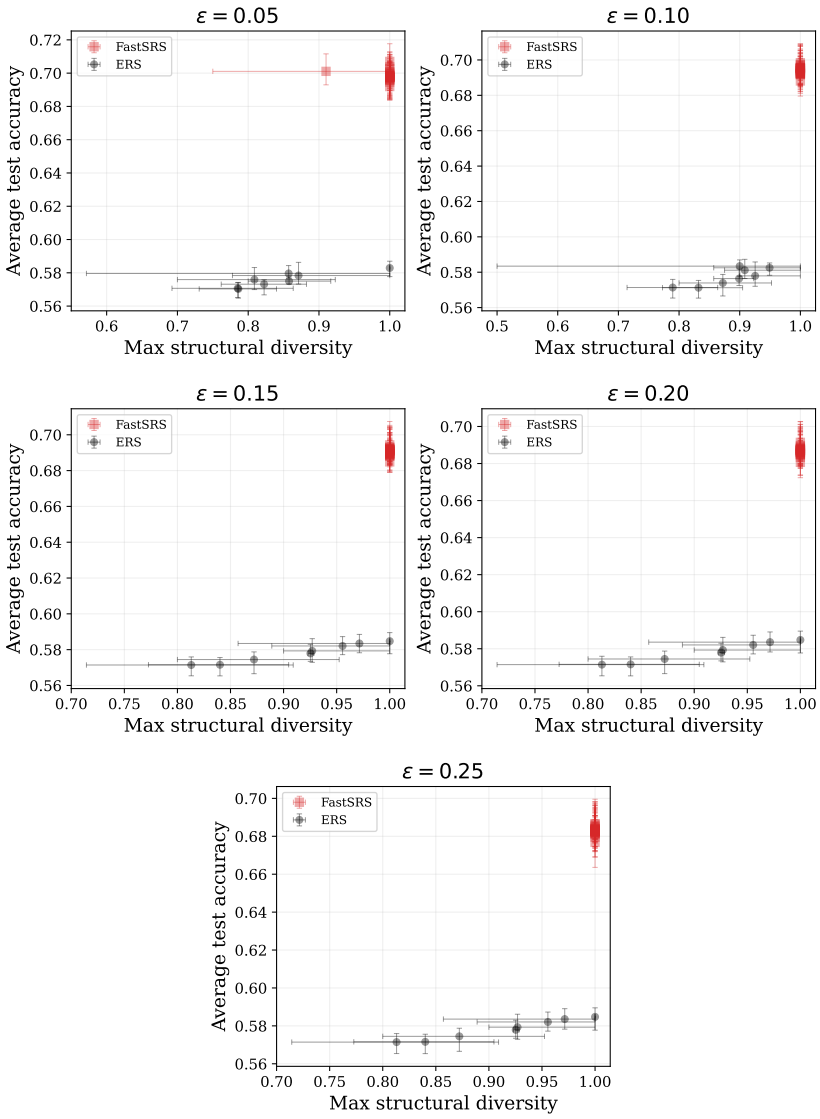

Fig. J47 Comparison between FastSRS and ERS in terms of test accuracy and maximum structural diversity for fico dataset.

fico

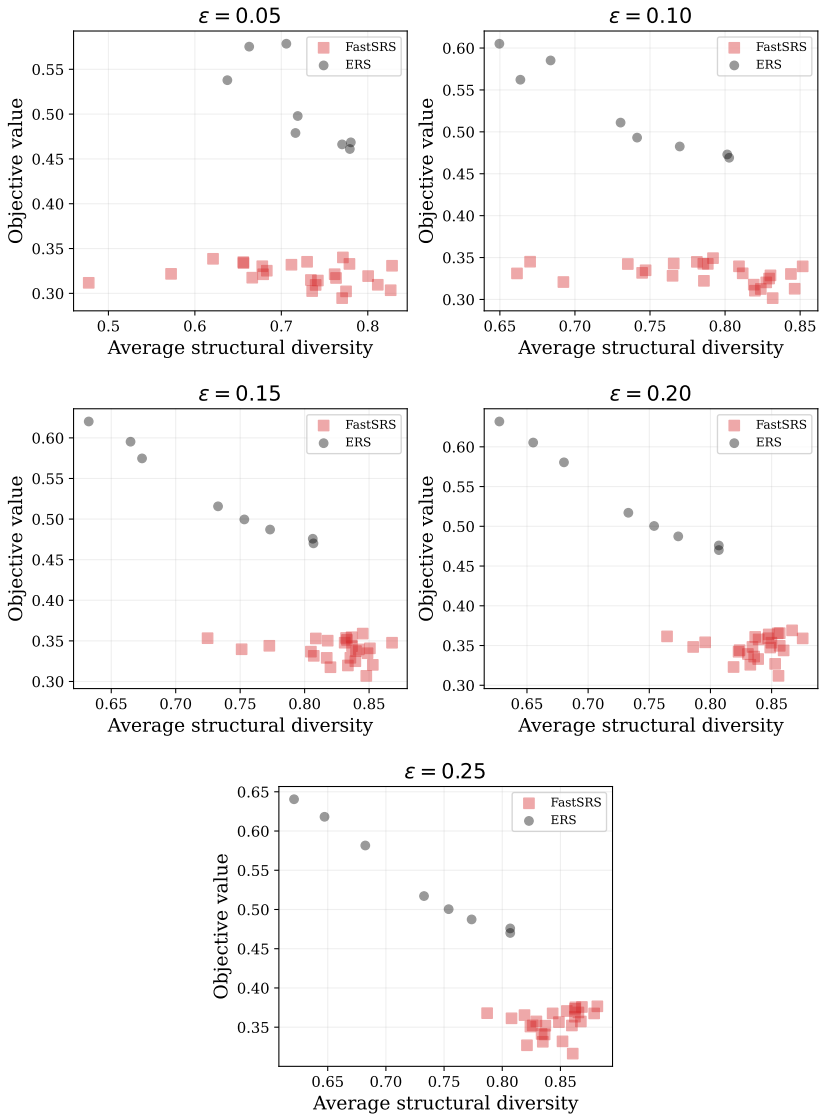

Fig. J48 Comparison between FastSRS and ERS in terms of objective value and average structural diversity for fico dataset.

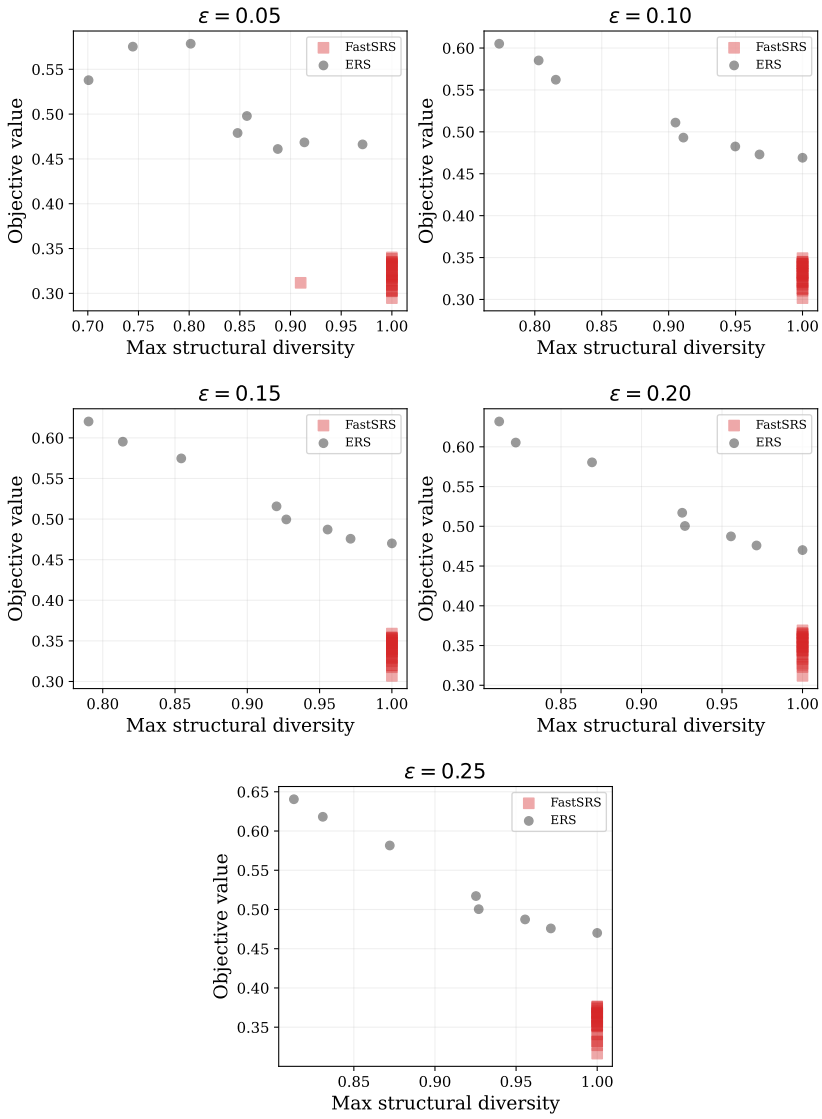

Fig. J49 Comparison between FastSRS and ERS in terms of objective value and maximum structural diversity for fico dataset.

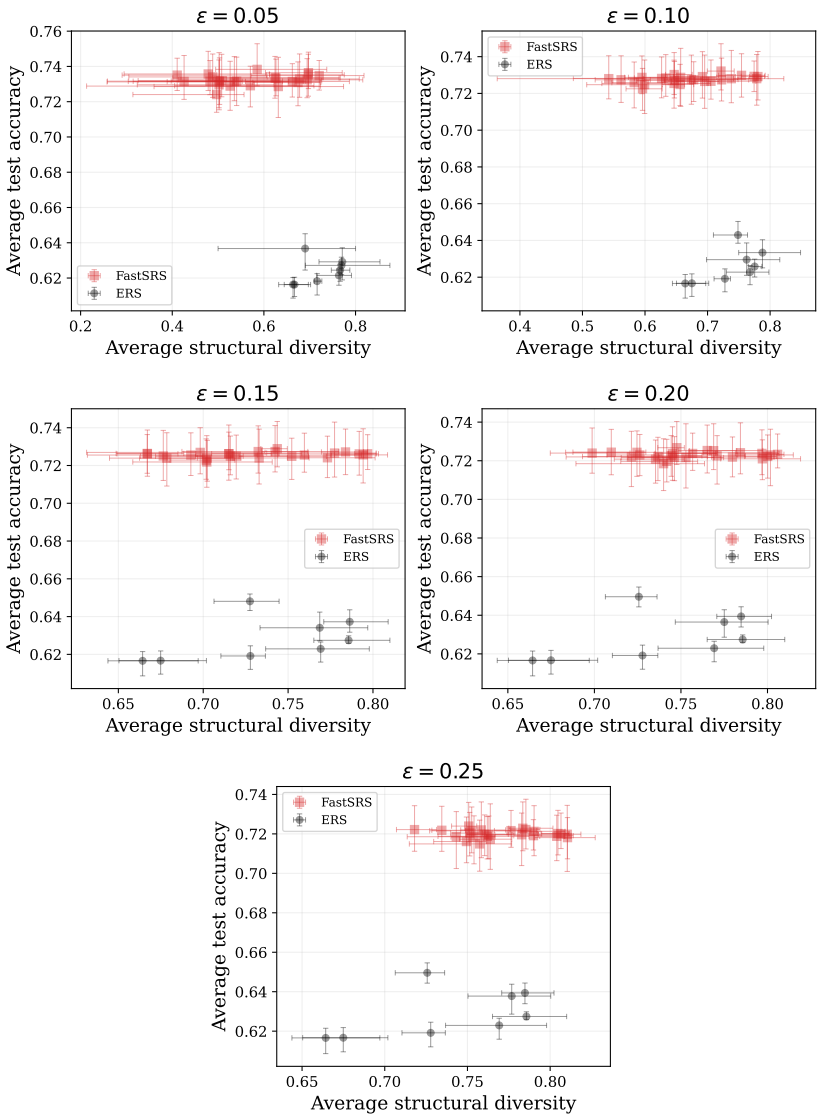

Fig. J50 Comparison between FastSRS and ERS in terms of test accuracy and average structural diversity for recidivism dataset.

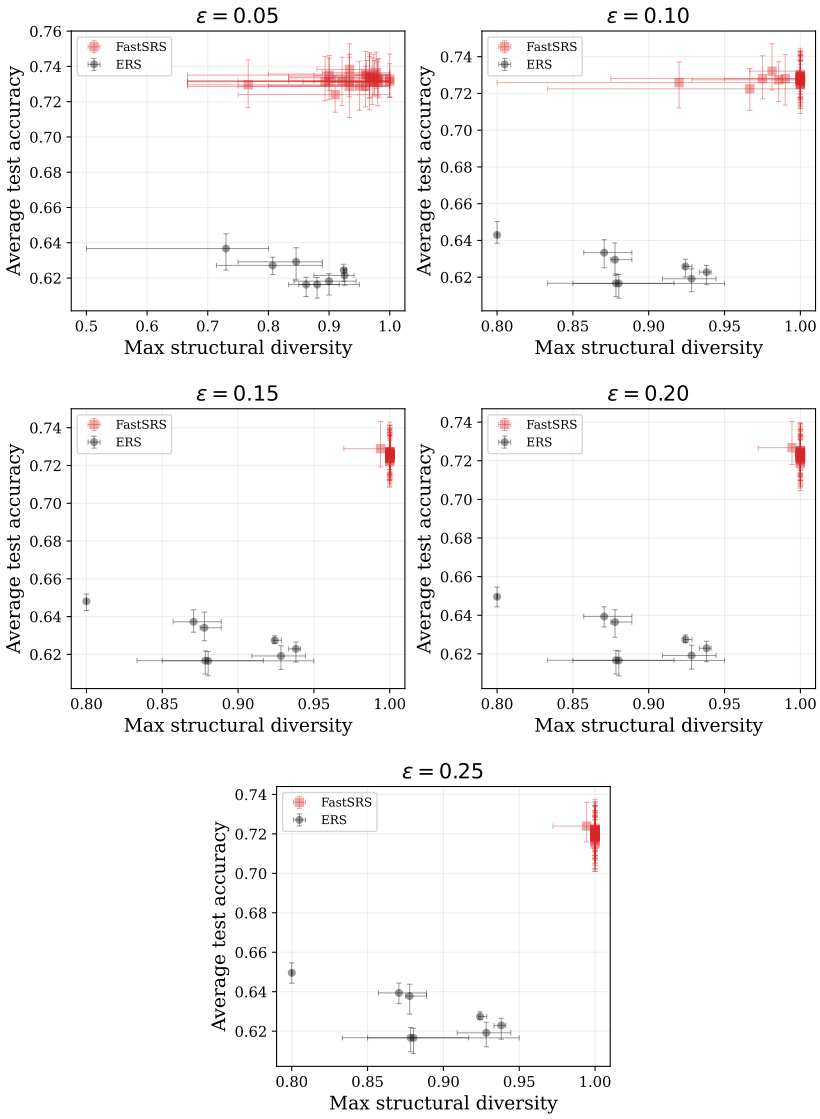

Fig. J51 Comparison between FastSRS and ERS in terms of test accuracy and maximum structural diversity for recidivism dataset.

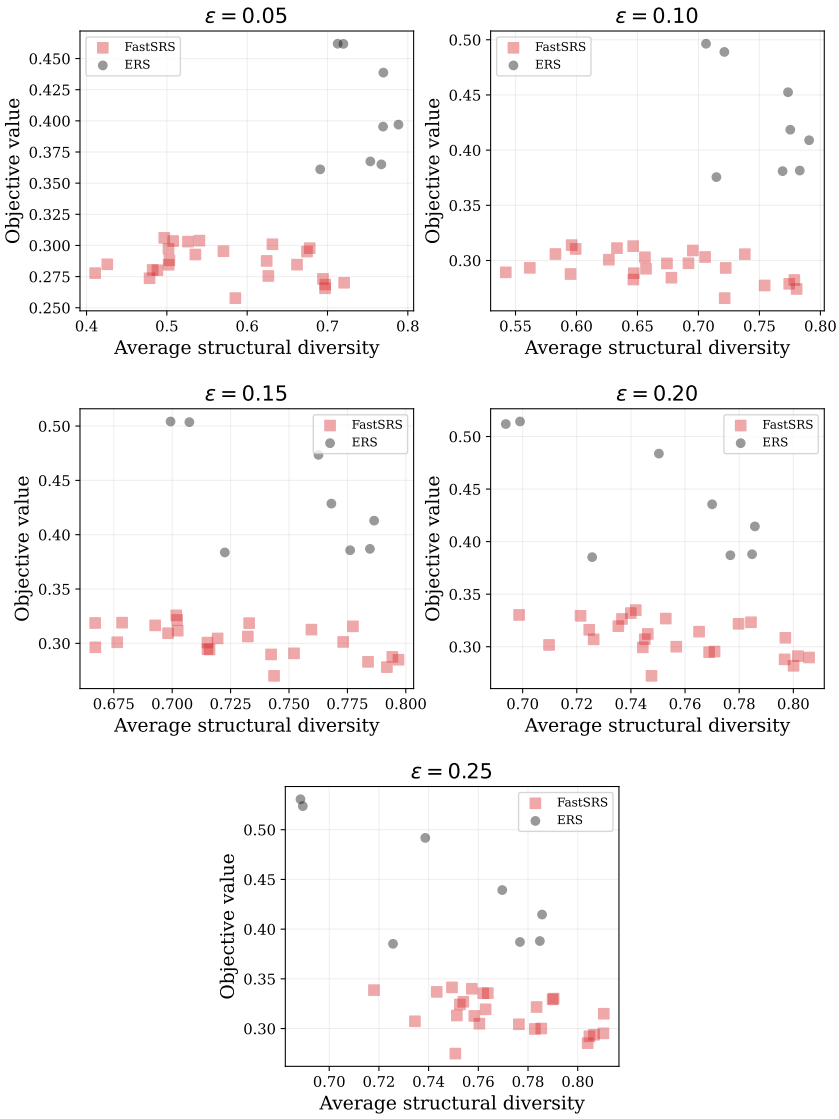

Fig. J52 Comparison between FastSRS and ERS in terms of objective value and average structural diversity for recidivism dataset.

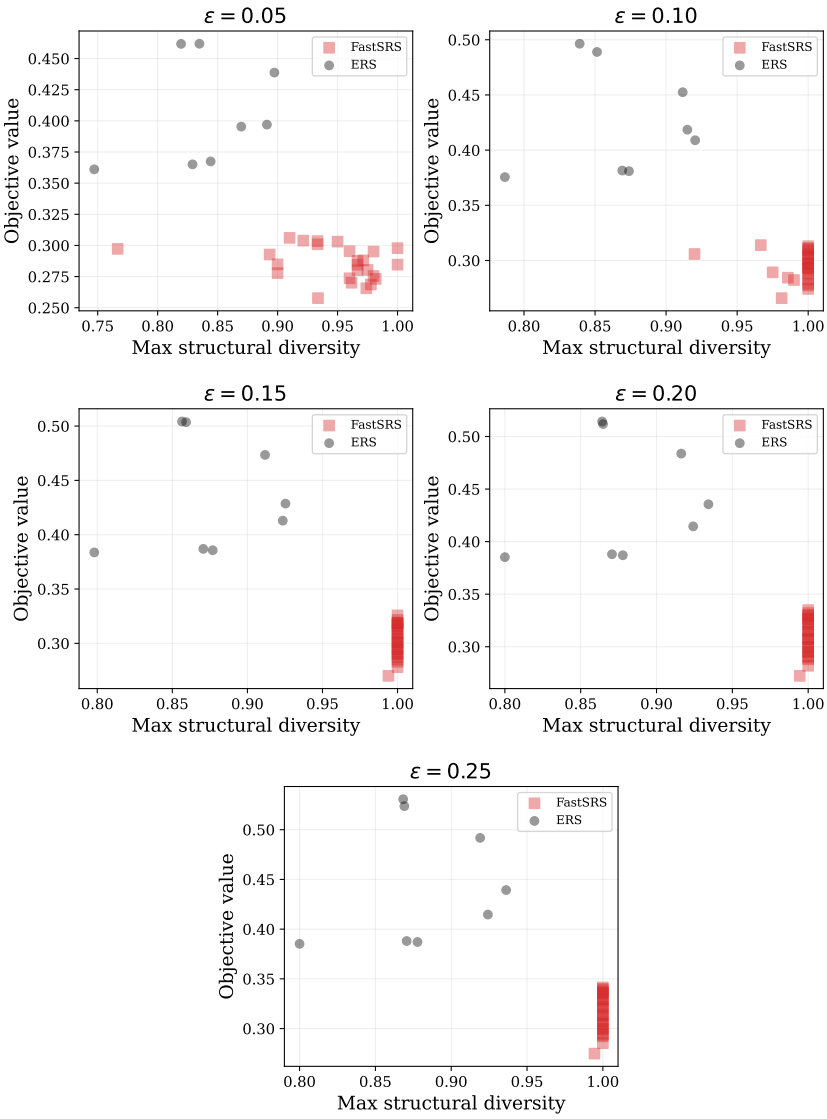

Fig. J53 Comparison between FastSRS and ERS in terms of objective value and maximum structural diversity for recidivism dataset.

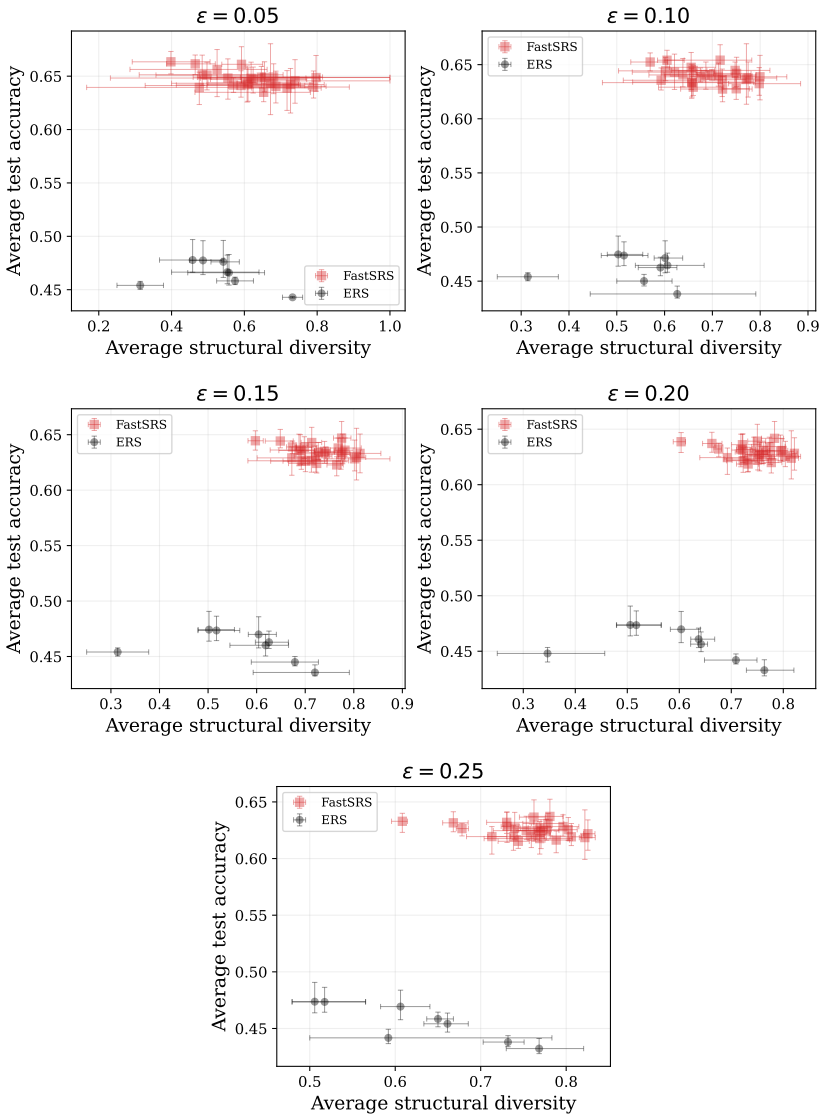

Fig. J54 Comparison between FastSRS and ERS in terms of test accuracy and average structural diversity for invehicle dataset.

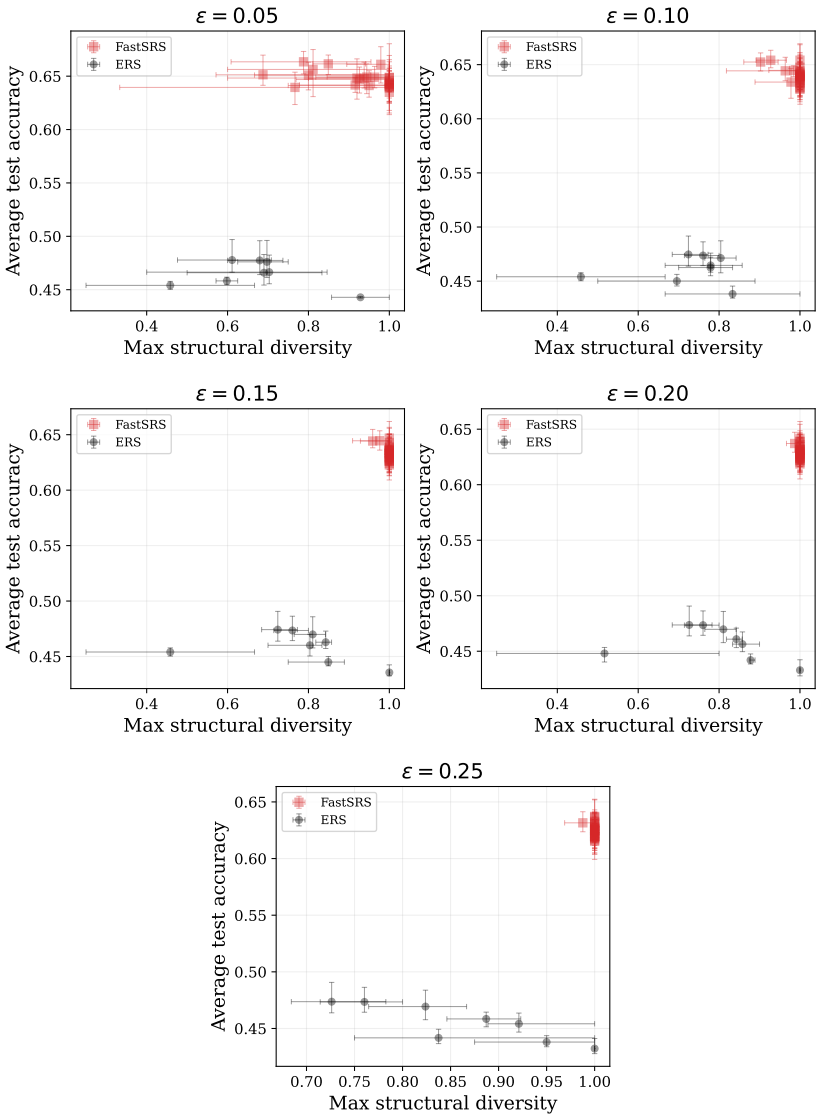

Fig. J55 Comparison between FastSRS and ERS in terms of test accuracy and maximum structural diversity for invehicle dataset.

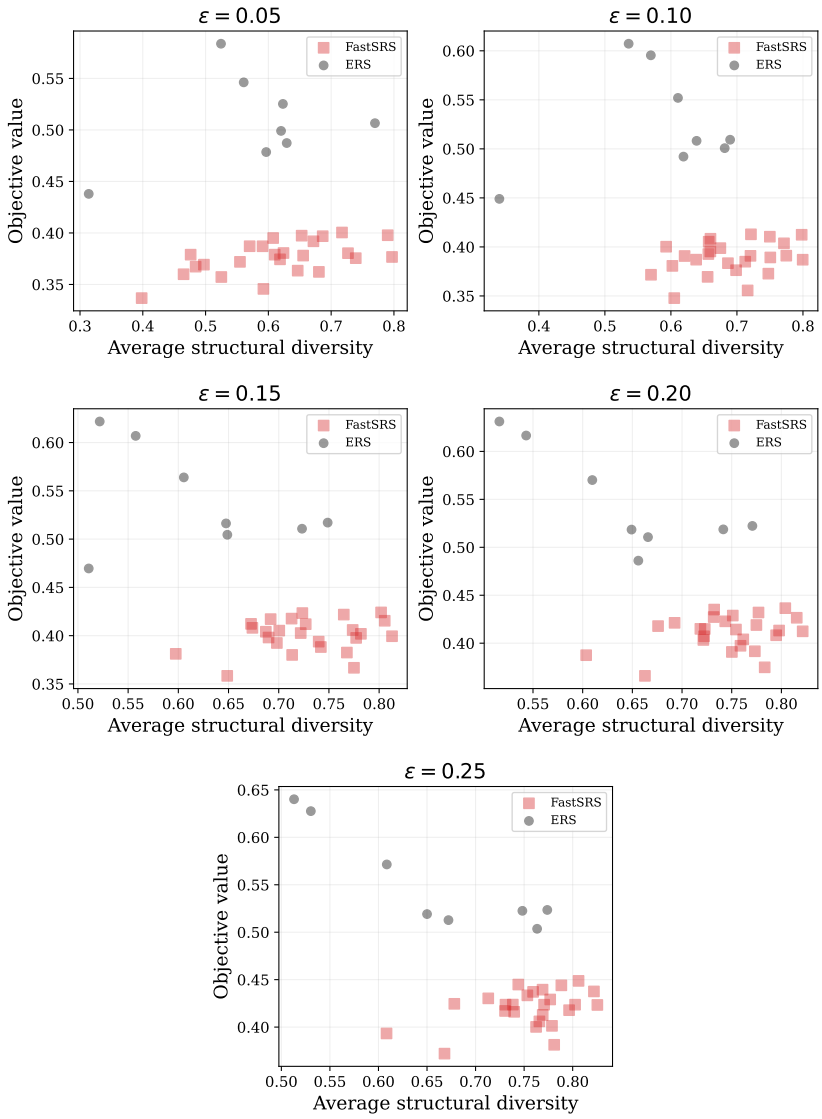

Fig. J56 Comparison between FastSRS and ERS in terms of objective value and average structural diversity for invehicle dataset.

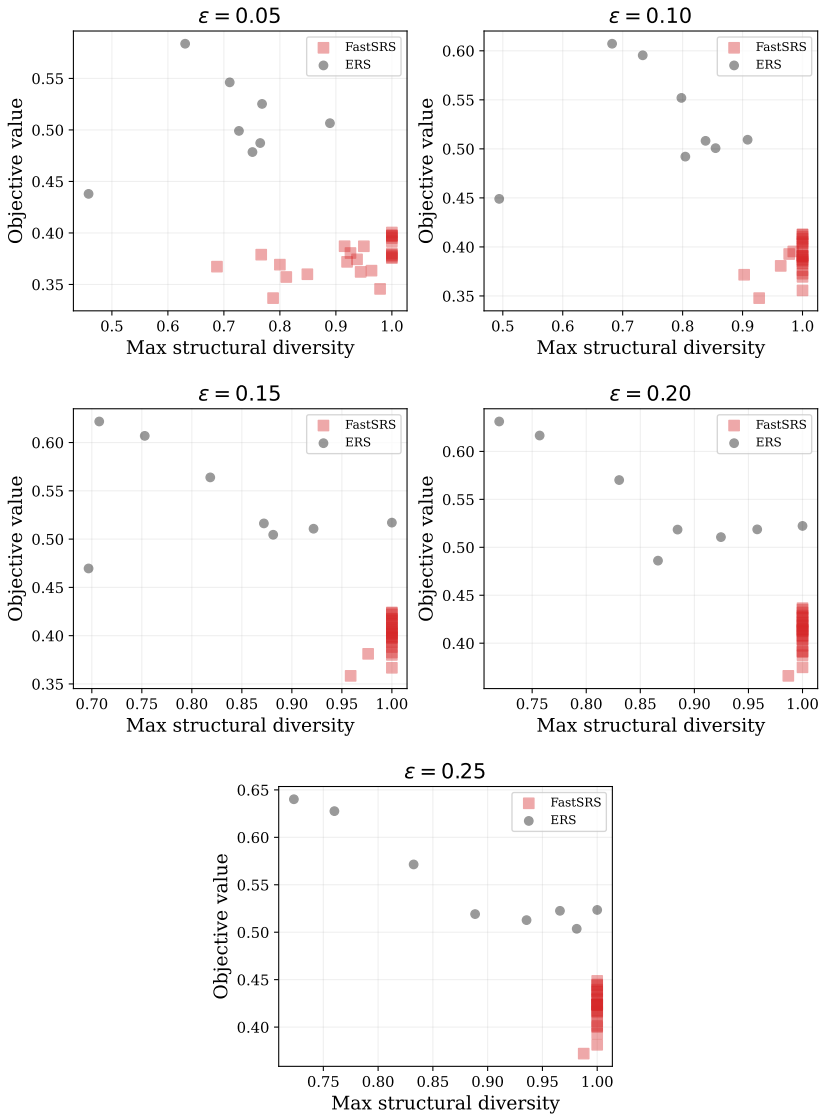

Fig. J57 Comparison between FastSRS and ERS in terms of objective value and maximum structural diversity for invehicle dataset.

adult

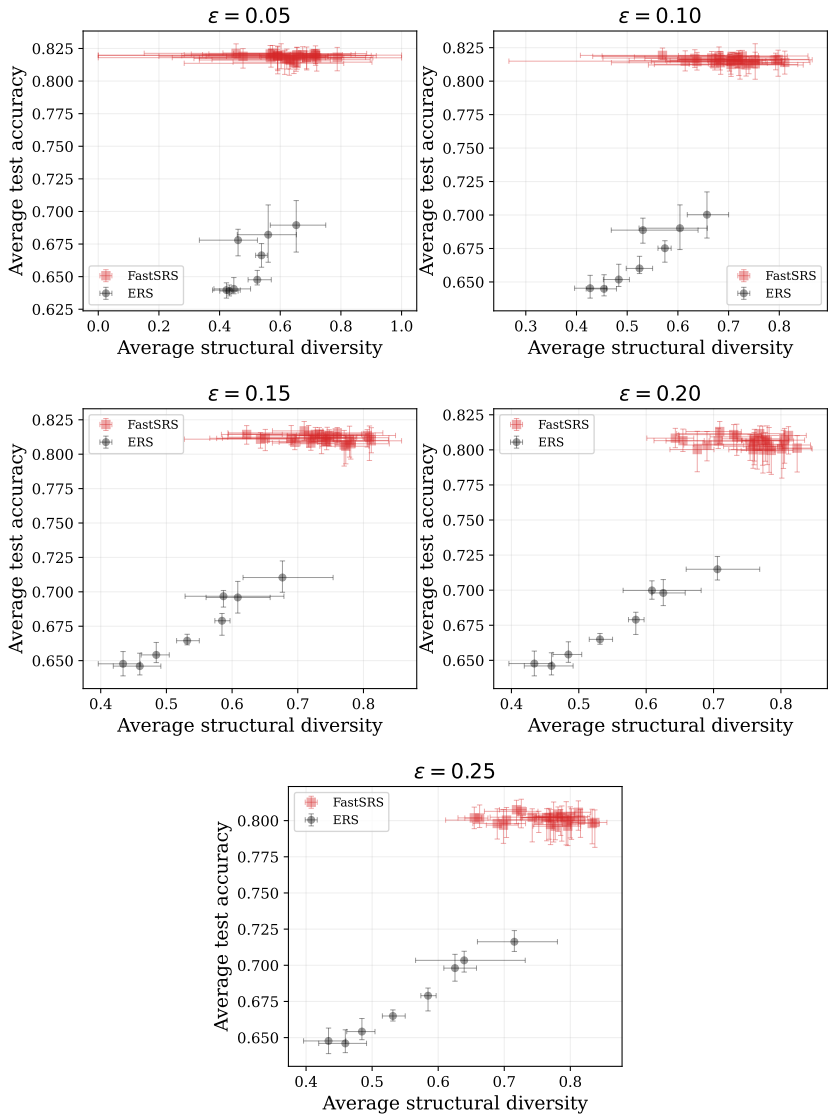

Fig. J58 Comparison between FastSRS and ERS in terms of test accuracy and average structural diversity for adult dataset.

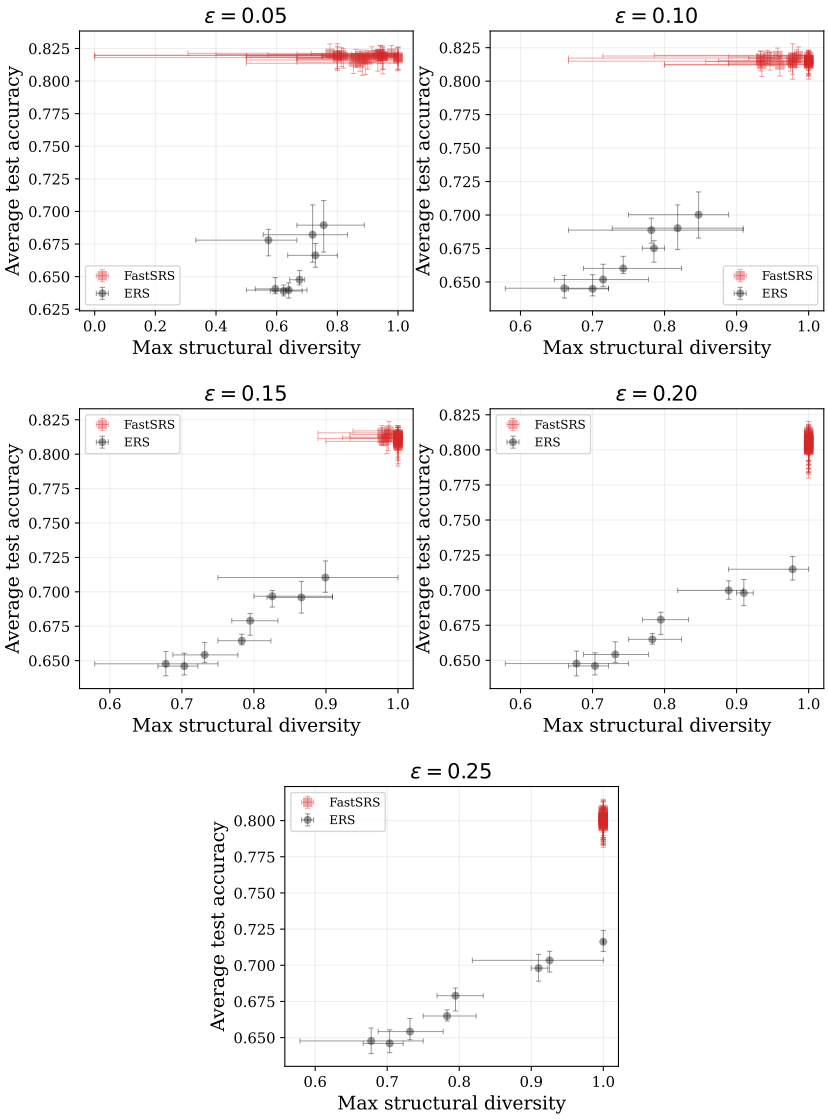

Fig. J59 Comparison between FastSRS and ERS in terms of test accuracy and maximum structural diversity for adult dataset.

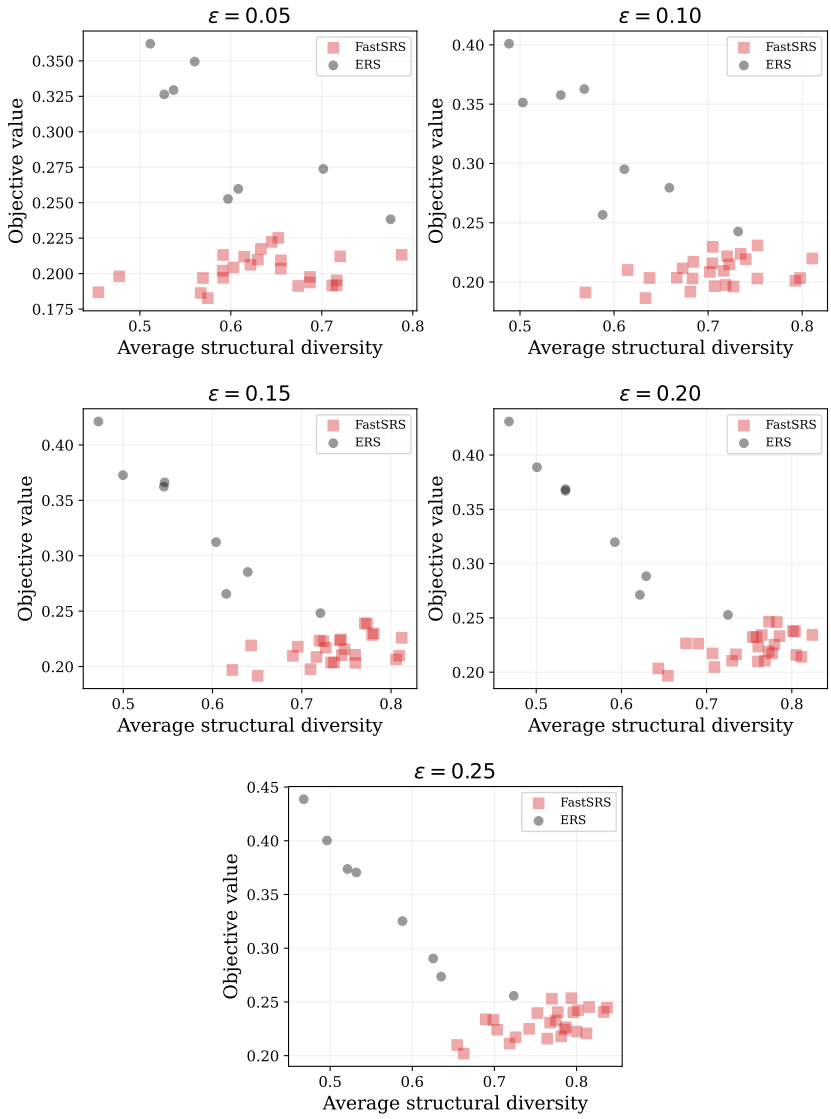

Fig. J60 Comparison between FastSRS and ERS in terms of objective value and average structural diversity for adult dataset.

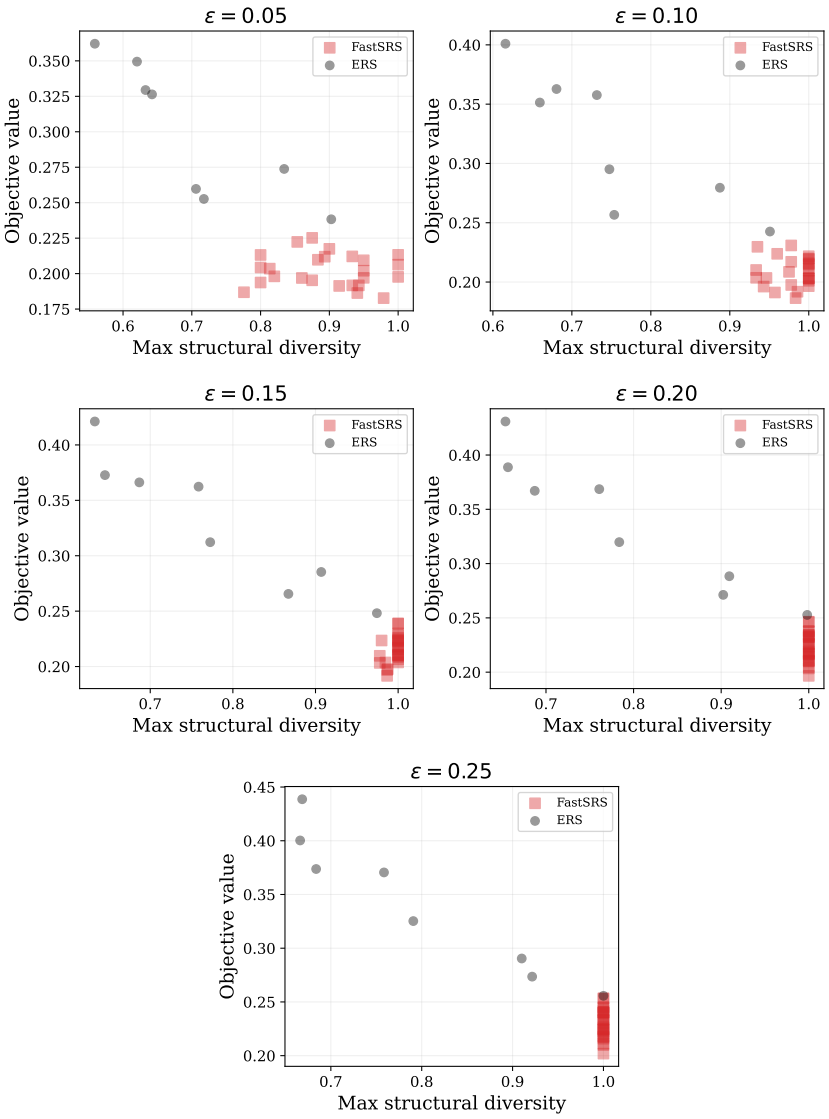

Fig. J61 Comparison between FastSRS and ERS in terms of objective value and maximum structural diversity for adult dataset.

Appendix K Approximate completeness of the Rashomon set produced by FastSRS

The true Rashomon set is generally intractable to compute exactly, as it would require enumerating all near-optimal rule sets. In order to assess empirical completeness, we have performed two experiments.

First, we study the fraction of the true Rashomon set captured by our sampled set for a synthetic dataset.

In particular, we construct a synthetic dataset with $J = 8$ binary features (16 with negations), $N = 1{,}000$ examples, and 5% label noise. The ground truth consists of 3 rules of maximum length 2. We exhaustively enumerate all 275,584,033 possible rule sets (up to 5 rules, each of length at most 2) and compute the true global optimum $R^* = \arg\min L(R)$ for each of 25 sparsity parameter configurations (C_2 ranging from 0.0005 to 0.025, with $C_1 = C_2/4$ and $C_3 = C_1/3$).

We measure coverage as the fraction of structurally distinct rule sets in the true Rashomon set that FastSRS exactly recovers. The coverage results, averaged across 25 parameter configurations, are summarized in Table K3.

ε	Distinct rule sets	Structural coverage	
	mean (max)	mean	median
0.05	1.5 (2)	78.0%	100.0%
0.10	2.4 (11)	74.0%	100.0%
0.15	7.3 (29)	61.2%	50.0%
0.20	14.7 (58)	49.7%	50.0%
0.25	102.0 (1,178)	32.5%	11.8%

Table K3 Coverage as a function of ϵ.

At $\varepsilon = 0.05$, FastSRS recovers 78% of the structurally distinct rule sets on average across 25 parameter configurations, achieving 100% coverage in more than half of the configurations. Coverage decreases for larger values of ε as the true Rashomon set grows rapidly (from at most 2 distinct rule sets at $\varepsilon = 0.05$ to 1,178 at $\varepsilon = 0.25$), while FastSRS stores at most 500 models per run — an inherent limitation of the iteration budget rather than the algorithm itself. At $\varepsilon = 0.25$, the Rashomon set can contain over a thousand rule sets, so 500 iterations cannot possibly discover them all. Running FastSRS with more iterations directly increases the number of explored models, and our scaling experiments (discussed below) confirm that the number of distinct rule sets discovered continues to grow with additional iterations.

Second, for real datasets, we assess empirical completeness by running FastSRS for a large number of iterations, namely, 50,000 iterations, in order to get the most complete Rashomon set we can, and then see how many iterations it takes to get to a fraction of it, as a function of ε. See Figure K62 below.

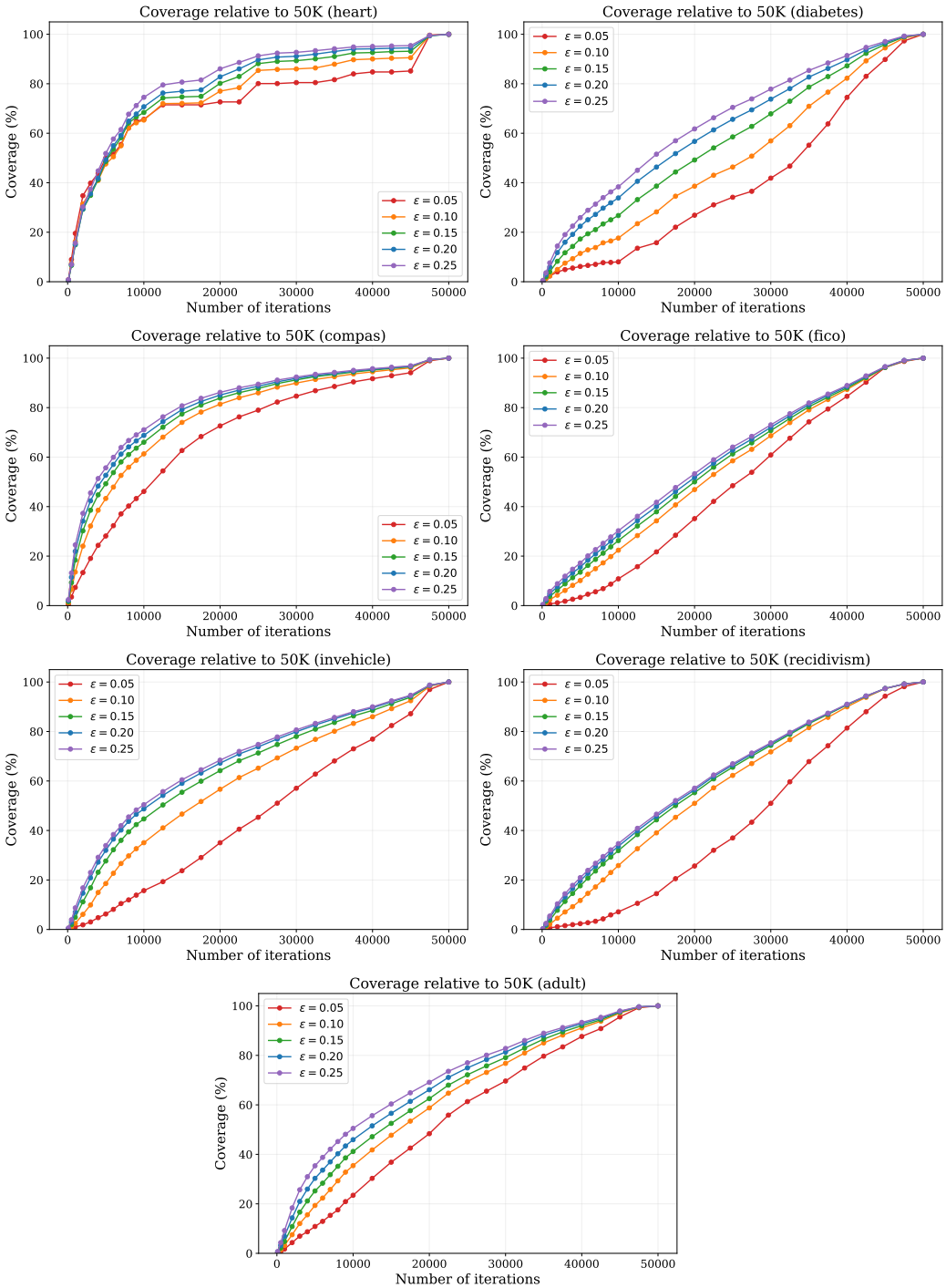

Fig. K62 Empirical completeness for real datasets.

Appendix L A collection of models from the Rashomon set produced by FastSRS for some (C_1, C_2, C_3) configuration

heart

Rule Set 0

```
rule 0:(cp:4.0 or 1.0),(ca:1.0 or not0.0),
rule 1:(thalach:<=146.4),(thal:7.0),
rule 2:(ca:not0.0),(slope:not3.0),(oldpeak:>1.9),
rule 3:(age:<=53.0),(thal:not3.0),(oldpeak:>1.5),
rule 4:(chol:>269.0),(thalach:<=159.0),(slope:not1.0),
```

Train Obj: 0.1496691 **Test Obj:** 0.2897535

Rule Set 3

```
rule 0:(cp:4.0 or 1.0),(ca:1.0 or not0.0),
rule 1:(thalach:<=146.4),(thal:7.0),
rule 2:(ca:not0.0),(oldpeak:>1.9),
rule 3:(chol:>269.0),(slope:not1.0),
rule 4:(age:<=53.0),(thal:not3.0),(oldpeak:>1.5),
```

Train Obj: 0.1581079 **Test Obj:** 0.2897535

Rule Set 4

```
rule 0:(cp:4.0 or 1.0),(ca:1.0 or not0.0),
rule 1:(thalach:<=146.4),(thal:7.0),
rule 2:(ca:not0.0),(oldpeak:>1.9),
rule 3:(age:<=66.0),(exang:1.0),(ca:2.0),
rule 4:(age:<=53.0),(thal:not3.0),(oldpeak:>1.5),
rule 5:(chol:>269.0),(thalach:<=159.0),(slope:not1.0),
```

Train Obj: 0.1496691 **Test Obj:** 0.2897535

Rule Set 8

```
rule 0:(cp:4.0 or 1.0),(ca:1.0 or not0.0),
rule 1:(thalach:<=146.4),(thal:7.0),
rule 2:(age:<=53.0),(thal:not3.0),(oldpeak:>1.5),
rule 3:(chol:>269.0),(thalach:<=159.0),(slope:not1.0),
```

Train Obj: 0.1581079 **Test Obj:** 0.3064201

Rule Set 6

```
rule 0:(thalach:<=146.4),(thal:7.0),
rule 1:(ca:not0.0),(oldpeak:>1.9),
rule 2:(cp:4.0 or 1.0),(ca:1.0 or not0.0),(exang:1.0),
```

```
rule 3:(age:<=53.0),(thal:not3.0),(oldpeak:>1.5),
rule 4:(chol:>269.0),(thalach:<=159.0),(slope:not1.0),
```

Train Obj: 0.1749855 **Test Obj:** 0.3230868

diabetes

Rule Set 0

```
rule 0:(Pregnancies:>7.0),(BMI:>41.5),
rule 1:(Glucose:>125.0),(BMI:>41.5),
rule 2:(DiabetesPedigreeFunction:>0.8786),(Pregnancies:>3.0),
rule 3:(BloodPressure:>54.0),(Glucose:>167.0),
rule 4:(Glucose:>134.0),(Pregnancies:>9.0),(SkinThickness:>27.0),
rule 5:(Glucose:>167.0 or >117.0),(BMI:>41.5 or >28.2),(BloodPressure:<=54.0),
```

Train Obj: 0.2372088 **Test Obj:** 0.2998384

Rule Set 2

```
rule 0:(Glucose:>167.0),
rule 1:(Pregnancies:>7.0),(BMI:>41.5),
rule 2:(Glucose:>125.0),(BMI:>41.5),
rule 3:(DiabetesPedigreeFunction:>0.8786),(Pregnancies:>3.0),
rule 4:(Glucose:>134.0),(Pregnancies:>9.0),(SkinThickness:>27.0),
rule 5:(Glucose: >117.0),(BMI:>28.2),(BloodPressure:<=54.0),
```

Train Obj: 0.2388375 **Test Obj:** 0.2998384

Rule Set 4

```
rule 0:(Pregnancies:>7.0),(BMI:>41.5),
rule 1:(Glucose:>125.0),(BMI:>41.5),
rule 2:(DiabetesPedigreeFunction:>0.8786),(Pregnancies:>3.0),
rule 3:(BloodPressure:>54.0),(Glucose:>167.0),
rule 4:(Age:>42.6),(Glucose:>117.0),
rule 5:(Glucose:>134.0),(Pregnancies:>9.0),(SkinThickness:>27.0),
rule 6:(Glucose:>117.0),(BMI: >28.2),(BloodPressure:<=54.0),
```

Train Obj: 0.2404662 **Test Obj:** 0.2673709

Rule Set 10

```
rule 0:(Pregnancies:>7.0),(BMI:>41.5),
rule 1:(Glucose:>125.0),(BMI:>41.5),
rule 2:(DiabetesPedigreeFunction:>0.8786),(Pregnancies:>3.0),
rule 3:(BloodPressure:>54.0),(Glucose:>167.0),
rule 4:(BMI:>28.2),(Glucose:>134.0),(BloodPressure:<=54.0),
rule 5:(Glucose:>134.0),(Pregnancies:>9.0),(SkinThickness:>27.0),
rule 6:(Glucose:>117.0),(BMI:>28.2),(BloodPressure:<=54.0),
```

Train Obj: 0.2372088 **Test Obj:** 0.2998384

Rule Set 14

```
rule 0:(Pregnancies:>7.0),(BMI:>41.5),
rule 1:(Glucose:>125.0),(BMI:>41.5),
rule 2:(BloodPressure:>54.0),(Glucose:>167.0),
rule 3:
(DiabetesPedigreeFunction:>0.8786),(Pregnancies:>3.0),(Glucose:<=109.0),
rule 4:(Glucose:>134.0),(Pregnancies:>9.0),(SkinThickness:>27.0),
rule 5:(Glucose: >117.0),(BMI:>28.2),(BloodPressure:<=54.0),
```

Train Obj: 0.2437235 **Test Obj:** 0.3063319

compas

Rule Set 0

```
rule 0:(priors.count:>2.0),(age:<=53.0),
rule 1:(age:<=22.0),(sex=female:not1),
```

Train Obj: 0.3436036 **Test Obj:** 0.3553646

Rule Set 3

```
rule 0:(priors.count:>2.0),(age:<=53.0),
rule 1:(age:<=31.0),(priors.count:>0.0),
```

Train Obj: 0.3597122 **Test Obj:** 0.3806902

Rule Set 7

```
rule 0:(priors.count:>2.0),(age:<=53.0),
```

Train Obj: 0.3627891 **Test Obj:** 0.3864789

Rule Set 8

```
rule 0:(priors.count:>2.0),(age:<=53.0),
rule 1:(age:<=22.0),(sex=female:not1),
rule 2:(current.charge.degree=felony:1),(priors.count:>1.0),
```

Train Obj: 0.3526533 **Test Obj:** 0.3618769

Rule Set 12

```
rule 0:(age:<=22.0),
rule 1:(priors.count:>2.0),(age:<=53.0),
```

Train Obj: 0.3463185 **Test Obj:** 0.3618769

fico

Rule Set 2

```
rule 0:(ExternalRiskEstimate:>75.0),
rule 1:(MSinceMostRecentInqexcl7days:>0.0),(ExternalRiskEstimate:>69.0),
rule 2:(ExternalRiskEstimate:>72.0),(NumSatisfactoryTrades:>26.0),
```

Train Obj: 0.29144296 **Test Obj:** 0.29033349

Rule Set 3

```
rule 0:(ExternalRiskEstimate:>75.0),
rule 1:(MSinceMostRecentInqexcl7days:>0.0),(ExternalRiskEstimate:>69.0),
rule 2:(ExternalRiskEstimate:>72.0),(NumSatisfactoryTrades:>26.0),
rule 3:(ExternalRiskEstimate:>72.0),(PercentTradesWBalance:<=60.0),
```

Train Obj: 0.29192103 **Test Obj:** 0.29272354

Rule Set 0

```
rule 0:(MSinceMostRecentInqexcl7days:>0.0),(ExternalRiskEstimate:>69.0),
rule 1:(AverageMInFile:>120.0),(MaxDelq2PublicRecLast12M:7),
rule 2:(NetFractionRevolvingBurden:<=12.0),(NumTotalTrades:>28.0),
```

Train Obj: 0.29383330 **Test Obj:** 0.30180576

Rule Set 1

```
rule 0:(ExternalRiskEstimate:>75.0),
rule 1:(MSinceMostRecentInqexcl7days:>0.0),(ExternalRiskEstimate:>69.0),
```

Train Obj: 0.29455041 **Test Obj:** 0.29750366

Rule Set 6

```
rule 0:(MSinceMostRecentInqexcl7days:>0.0),(ExternalRiskEstimate:>69.0),
rule 1:(ExternalRiskEstimate:>72.0),(NumSatisfactoryTrades:>26.0),
rule 2:
(AverageMInFile:>92.0),(ExternalRiskEstimate:>78.0),(NumTotalTrades:>40.0),
```

Train Obj: 0.32968845 **Test Obj:** 0.33813463

adult

Rule Set 0

```
rule 0:(education.num:>10.0),(capital.gain:>0.0),
rule 1:(education.num:>10.0),(marital.status: Married.civ.spouse),
rule 2:(education.num:>9.0),(marital.status:
Married.civ.spouse),(capital.gain:>0.0),
```

Train Obj: 0.1796476 **Test Obj:** 0.1781618

Rule Set 1

```
rule 0:(education.num:>10.0),(capital.gain:>0.0),
rule 1:(education.num:>10.0),(marital.status: Married.civ.spouse),
rule 2:(education.num:>9.0 or >7.0),(marital.status: Married.civ.spouse),
(capital.gain:>0.0),
```

Train Obj: 0.1773058 **Test Obj:** 0.1781618

Rule Set 7

```
rule 0:(education.num:>10.0),(capital.gain:>0.0),
rule 1:(education.num:>7.0),(relationship: Wife),
rule 2:(education.num:>7.0),(marital.status:
Married.civ.spouse),(capital.gain:>0.0),
```

Train Obj: 0.2162723 **Test Obj:** 0.2137829

Rule Set 8

```
rule 0:(education.num:>10.0),(capital.gain:>0.0),
rule 1:(age:>41.0),(education.num:>13.0),(workclass: Self.emp.inc),
rule 2:(education.num:>7.0),(marital.status:
Married.civ.spouse),(capital.gain:>0.0),
```

Train Obj: 0.2135082 **Test Obj:** 0.2044170

Rule Set 10

```
rule 0:(education.num:>10.0),(capital.gain:>0.0),
rule 1:(education.num:>10.0),(marital.status:
Married.civ.spouse),(capital.loss:>0.0),
rule 2:(education.num:>7.0),(marital.status:
Married.civ.spouse),(capital.gain:>0.0),
```

Train Obj: 0.2039105 **Test Obj:** 0.1952047

invehicle

Rule Set 0

```
rule 0:(coupon:Carry out & Take away),
rule 1:(coupon:Restaurant(<20)),(expiration:not2h),
rule 2:(expiration:not2h),(passanger:Friend(s)),
rule 3:(maritalStatus:Single),(RestaurantLessThan20:gt8),
rule 4:(coupon:notBar),(destination:No Urgent Place),
rule 5:(CoffeeHouse:1~3),(toCoupon.GEQ15min:0),
```

Train Obj: 0.3475963 **Test Obj:** 0.3519070

Rule Set 2

```
rule 0:(coupon:Carry out & Take away),
rule 1:(coupon:Restaurant(<20)),(expiration:not2h),
rule 2:(coupon:notBar),(destination:No Urgent Place),
rule 3:(CoffeeHouse:1~3),(toCoupon.GEQ15min:0),
rule 4:(direction.same:not0),(time:6PM),
rule 5:(Bar:1~3),(expiration:not2h),
rule 6:(maritalStatus:Single),(RestaurantLessThan20:gt8),
(occupation:Computer & Mathematical),
```

Train Obj: 0.3435602 **Test Obj:** 0.3494236

Rule Set 8

```
rule 0:(coupon:Carry out & Take away),
rule 1:(coupon:Restaurant(<20)),(expiration:not2h),
rule 2:(CoffeeHouse:1~3),(toCoupon.GEQ15min:0),
rule 3:(Bar:1~3),(expiration:not2h),
rule 4:(CoffeeHouse:notnever),(passanger:Friend(s) or Partner),
rule 5:(coupon:Restaurant(<20)),(destination:No Urgent Place),
rule 6:(coupon:notBar),(time:10AM),
rule 7:(RestaurantLessThan20:gt8),(occupation:Computer &
Mathematical),(age:21),
```

Train Obj: 0.3283476 **Test Obj:** 0.3345229

Rule Set 10

```
rule 0:(coupon:Carry out & Take away or Restaurant(<20)),
rule 1:(CoffeeHouse:1~3),(toCoupon.GEQ15min:0),
rule 2:(Bar:1~3),(expiration:not2h),
rule 3:(CoffeeHouse:notnever),(passanger:Friend(s) or Partner),
rule 4:(coupon:notBar),(time:10AM),
rule 5:(RestaurantLessThan20:gt8),(occupation:Computer &
Mathematical),(age:21),
```

Train Obj: 0.3374545 **Test Obj:** 0.3452845

Rule Set 12

```
rule 0:(coupon:Carry out & Take away),
rule 1:(coupon:Restaurant(<20)),(expiration:not2h),
rule 2:(CoffeeHouse:1~3),(toCoupon.GEQ15min:0),
rule 3:(Bar:1~3),(expiration:not2h),
rule 4:(coupon:Restaurant(<20)),(destination:No Urgent Place),
rule 5:(coupon:notBar),(time:10AM),
rule 6:(CoffeeHouse:notnever),(passanger:Friend(s) or Partner),
       (RestaurantLessThan20:never),
rule 7:(RestaurantLessThan20:gt8),(occupation:Computer &
Mathematical),(age:21),
```

Train Obj: 0.3385928 **Test Obj:** 0.3378342

recidivism

Rule Set 0

```
rule 0:(Reasonforbeingtransferred:Prison),
rule 1:(Weightedtimeprobationtofirstviolation.days:1 or
2),(Jurisdiction:Philadelphia,PA),
rule 2:(Weightedtimeprobationtofirstviolation.days:1 or 0),
       (Distributionofsupervisionfees:$500to$999),
rule 3:(Arrestdatematch.rapsheetandsentencingsurvey:No),
       (Recodeoftimefromviolationtohearing:63ormoredays),
rule 4:
(Recodeoftimefromviolationtohearing:63ormoredays),(Stillonprobation:Yes),
rule 5:(Drugabusehistory:Frequentabuse),(Recodeoftimefromviolationtohearing:6
3ormoredays),
rule 6:(Recodeoftimefromviolationtohearing:15to31days),(Jurisdiction:SanFranc
isco,CA),
rule 7:(Probationrecordsreflectjailbeingimposed:Yes),
       (Recodeoftimetofirsthearing:549ormoredays),
       (Distributionofpercentofassessments.other:1to24%),
rule 8:(Reasonforbeingtransferred:Prison or Notascertained),
       (Threedigitoffensecode:Escape or
       Burglary),(Jurisdiction:DadeCounty,FL),
rule 9:(Probationfeesimposed:Yes,amountknown),(Recodeoftimetofirsthearing:184
to365days),
       (1986convictionoffense:Robbery),
rule 10:(Recodeoftimefromviolationtohearing:63ormoredays or 15to31days),
        (Stillonprobation:Yes or
        No),(Threedigitoffensecode:Robbery),(Age.weighted:0),
```

Train Obj: 0.2624901 **Test Obj:** 0.2557275

Rule Set 4

```
rule 0:(Reasonforbeingtransferred:Prison),
rule 1:(Weightedtimeprobationtofirstviolation.days:1 or
2),(Jurisdiction:Philadelphia,PA),
rule 2:(Weightedtimeprobationtofirstviolation.days:1 or 0),
       (Distributionofsupervisionfees:$500to$999),
rule 3:(Arrestdatematch.rapsheetandsentencingsurvey:No),
       (Recodeoftimefromviolationtohearing:63ormoredays),
rule 4:
(Recodeoftimefromviolationtohearing:63ormoredays),(Stillonprobation:Yes),
rule 5:(Drugabusehistory:Frequentabuse),(Recodeoftimefromviolationtohearing:6
3ormoredays),
rule 6:(Recodeoftimefromviolationtohearing:15to31days),(Jurisdiction:SanFranc
isco,CA),
```

```
rule 7:(Recodeoftimefromviolationtohearing:63ormoredays),(Weightedrisk:5 or
4),
rule 8:
(Reasonforbeingtransferred:Notascertained),(Jurisdiction:DadeCounty,FL),
       (Threedigitoffensecode:Escape or Burglary),
rule 9:(Probationfeesimposed:Yes,amountknown),(Recodeoftimetofirsthearing:184
to365days),
       (1986convictionoffense:Robbery),
rule 10:(Recodeoftimefromviolationtohearing:63ormoredays or 15to31days),
        (Stillonprobation:Yes or
        No),(Threedigitoffensecode:Robbery),(Age.weighted:0),
```

Train Obj: 0.2616314 **Test Obj:** 0.2561569

Rule Set 6

```
rule 0:(Reasonforbeingtransferred:Prison),
rule 1:(Weightedtimeprobationtofirstviolation.days:1 or 2),
(Jurisdiction:Philadelphia,PA),
rule 2:(Weightedtimeprobationtofirstviolation.days:1 or 0),
       (Distributionofsupervisionfees:$500to$999),
rule 3:(Arrestdatematch.rapsheetandsentencingsurvey:No),
       (Recodeoftimefromviolationtohearing:63ormoredays),
rule 4:
(Recodeoftimefromviolationtohearing:63ormoredays),(Stillonprobation:Yes),
rule 5:(Drugabusehistory:Frequentabuse),
(Recodeoftimefromviolationtohearing:63ormoredays),
rule 6:(Recodeoftimefromviolationtohearing:15to31days),(Jurisdiction:SanFranc
isco,CA),
rule 7:(Recodeoftimefromviolationtohearing:63ormoredays),(Weightedrisk:5 or
4),
rule 8:(Threedigitoffensecode:Escape or
Burglary),(Jurisdiction:DadeCounty,FL),
rule 9:(Arrestdatematch.rapsheetandsentencingsurvey:No),
    (Distributionofallotherfees:$100to$249),
rule 10:(Probationfeesimposed:Yes,amountknown),
(Recodeoftimetofirsthearing:184to365days),
      (1986convictionoffense:Robbery),
rule 11:(Recodeoftimefromviolationtohearing:63ormoredays or 15to31days),
      (Stillonprobation:Yes or
      No),(Threedigitoffensecode:Robbery),(Age.weighted:0),
```

Train Obj: 0.2588405 **Test Obj:** 0.2535807

Rule Set 9

```
rule 0:(Reasonforbeingtransferred:Prison),
rule 1:(Weightedtimeprobationtofirstviolation.days:1 or 2),
       (Jurisdiction:Philadelphia,PA),
rule 2:(Weightedtimeprobationtofirstviolation.days:1 or 0),
       (Distributionofsupervisionfees:$500to$999),
```

```
rule 3:(Arrestdatematch.rapsheetandsentencingsurvey:No),
       (Recodeoftimefromviolationtohearing:63ormoredays),
rule 4:
(Recodeoftimefromviolationtohearing:63ormoredays),(Stillonprobation:Yes),
rule 5:(Drugabusehistory:Frequentabuse),
       (Recodeoftimefromviolationtohearing:63ormoredays),
rule 6:(Recodeoftimefromviolationtohearing:15to31days),(Jurisdiction:SanFranc
isco,CA),
rule 7:(Recodeoftimefromviolationtohearing:63ormoredays),(Weightedrisk:5 or
4),
rule 8:(Threedigitoffensecode:Escape or
Burglary),(Jurisdiction:DadeCounty,FL),
rule 9:(Arrestdatematch.rapsheetandsentencingsurvey:No),
       (Distributionofallotherfees:$100to$249),
rule 10:(Lastsupervisionlevel:Maximumsupervision),
(Recodeoftimetofirsthearing:549ormoredays),
rule 11:(Probationfeesimposed:Yes,amountknown),
(Recodeoftimetofirsthearing:184to365days),
        (1986convictionoffense:Robbery),
rule 12:(Recodeoftimefromviolationtohearing:63ormoredays or 15to31days),
        (Stillonprobation:Yes or
        No),(Threedigitoffensecode:Robbery),(Age.weighted:0),
```

Train Obj: 0.2578744 **Test Obj:** 0.2527219

Rule Set 14

```
rule 0:(Reasonforbeingtransferred:Prison),
rule 1:(Weightedtimeprobationtofirstviolation.days:1 or
2),(Jurisdiction:Philadelphia,PA),
rule 2:(Weightedtimeprobationtofirstviolation.days:1 or 0),
       (Distributionofsupervisionfees:$500to$999),
rule 3:(Arrestdatematch.rapsheetandsentencingsurvey:No),
       (Recodeoftimefromviolationtohearing:63ormoredays),
rule 4:
(Recodeoftimefromviolationtohearing:63ormoredays),(Stillonprobation:Yes),
rule 5:(Drugabusehistory:Frequentabuse),(Recodeoftimefromviolationtohearing:6
3ormoredays),
rule 6:(Recodeoftimefromviolationtohearing:15to31days),(Jurisdiction:SanFranc
isco,CA),
rule 7:(Recodeoftimefromviolationtohearing:63ormoredays),(Weightedrisk:5 or
4),
rule 8:(Threedigitoffensecode:Escape or
Burglary),(Jurisdiction:DadeCounty,FL),
rule 9:(Lastsupervisionlevel:Maximumsupervision),(Recodeoftimetofirsthearing:
549ormoredays),
rule 10:(Raceofprobationer:Black),(Transferredtoanotherjurisdiction:Notascert
ained),
rule 11:(Probationfeesimposed:Yes,amountknown),(Recodeoftimetofirsthearing:18
4to365days),
```

```
rule 12:(Recodeoftimefromviolationtohearing:63ormoredays or 15to31days),
        (Stillonprobation:Yes or
        No),(Threedigitoffensecode:Robbery),(Age.weighted:0),
```

Train Obj: 0.2605579 **Test Obj:** 0.2561569

Appendix M Effectiveness of bounds in Theorems 7, 8 and 9

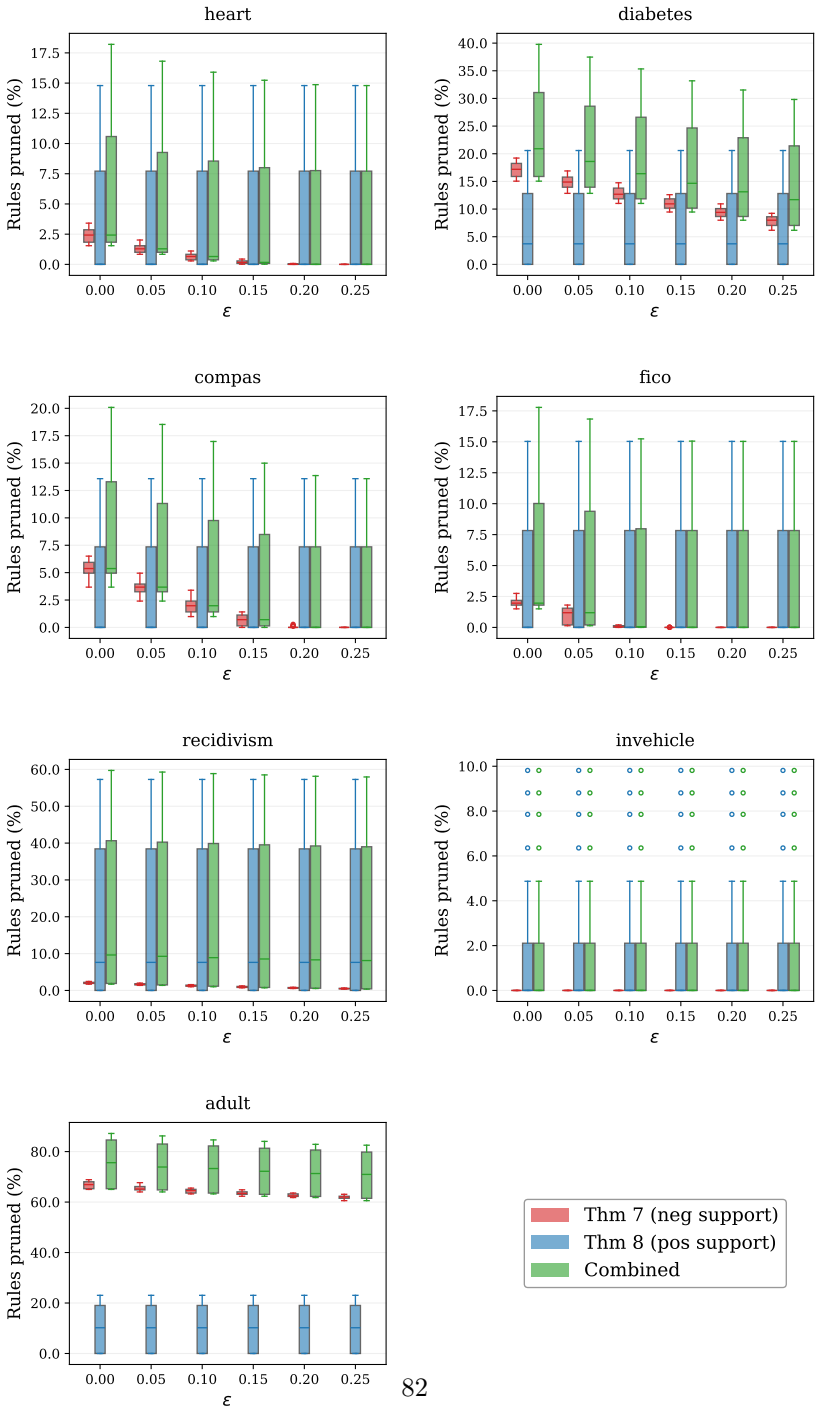

Fig. M63 Average distribution across sparsity parameters of the proportion of candidate rules pruned by Theorems 7 and 8 as a function of ε.

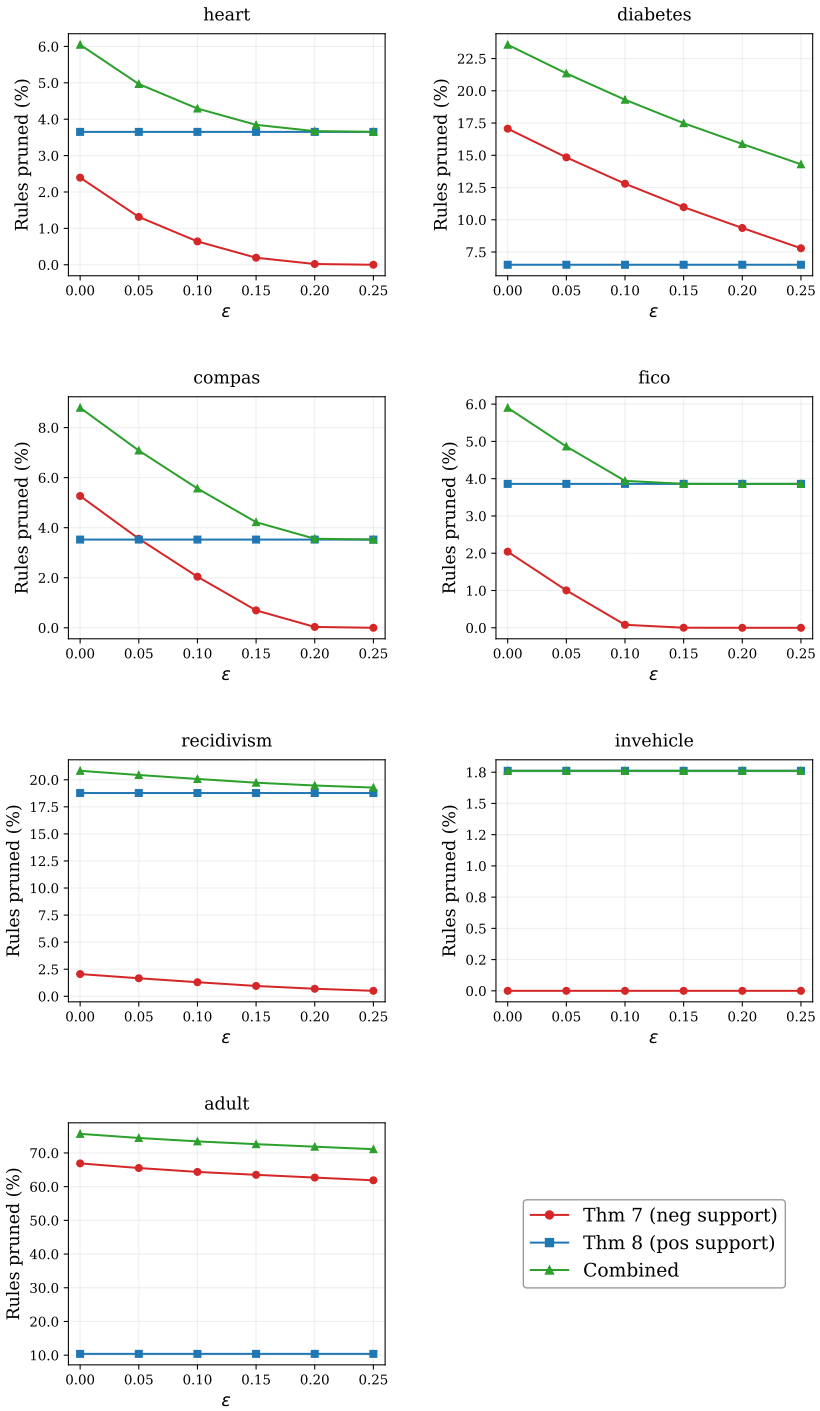

Fig. M64 Average proportion of candidate rules pruned by Theorems 7 and 8 as a function of ε.

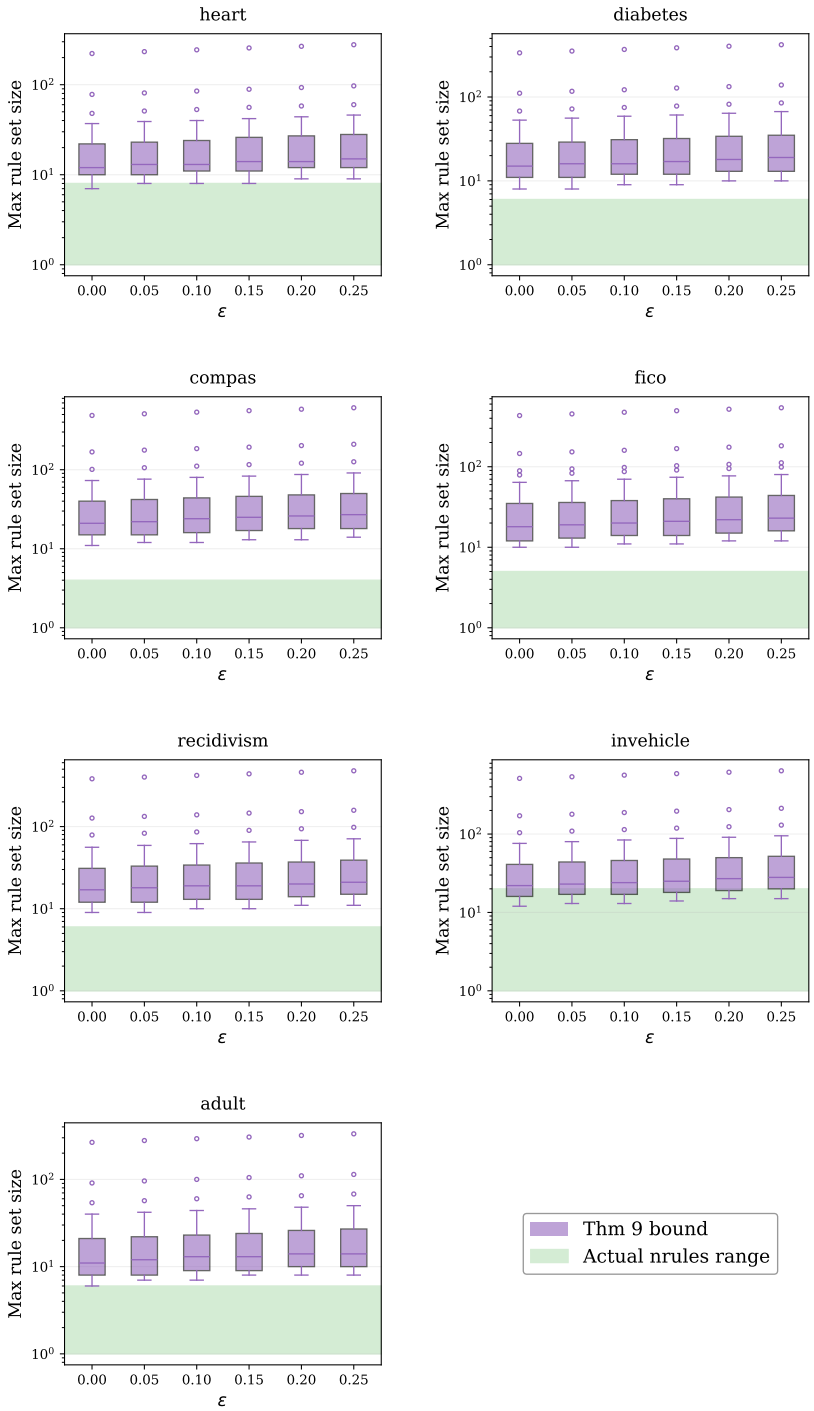

Fig. M65 Dynamic bound derived by Theorem 9 as a function of ε. The purple boxes show the distribution of Theorem 9 max rule set size across 25 C values; the green band shows the range of the actual number of rules found by FastSRS. The bound is effective in helping to reduce the search space even when it is above the green band; however when the lower whiskers of the boxes (minimum bound, around 6–12 rules) meet the green band, these are cases where Theorem 9 is fairly tight and likely to be eliminating large parts of the search space for some C configurations. See, for instance, heart, invehicle and adult. As expected, the bounds become progressively looser with ε.

Appendix N Scalability of FastSRS

To assess how runtime scales with dataset size for FastSRS, we used the synthetic dataset described in Appendix E, with $p = 100$ and varying the dataset size in $N \in \{100, 500, 1000, 5000, 10000, 50000, 100000\}$. The corresponding results are reported below in Figure N66. The observed runtime trend increases smoothly and regularly with dataset size across the evaluated range. This behavior is empirically consistent with a polynomial growth pattern, with both variants exhibiting similar scaling trends and differing mainly by constant factors.

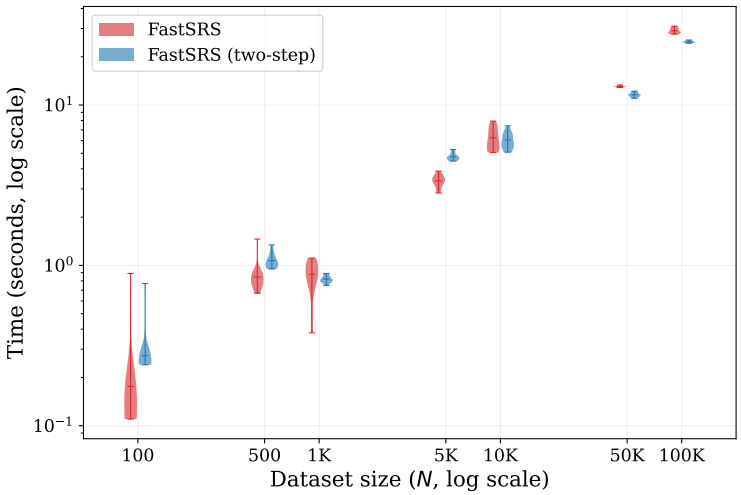

Fig. N66 Runtime as a function of dataset size

References

Akyüz, M.H., Birbil, Ş.İ.: Discovering classification rules for interpretable learning with linear programming. arXiv preprint arXiv:2104.10751 (2021)

Alcalá-Fdez, J., Alcalá, R., Herrera, F.: A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning. IEEE Transactions on Fuzzy Systems **19**(5), 857–872 (2011)

Angelino, E., Larus-Stone, N., Alabi, D., Seltzer, M., Rudin, C.: Learning certifiably optimal rule lists for categorical data. Journal of Machine Learning Research **18**(234), 1–78 (2018)

Bénard, C., Biau, G., Veiga, S.D., Scornet, E.: SIRUS: Stable and Interpretable RUle Set for classification. Electronic Journal of Statistics **15**, 427–505 (2021)

Cohen, W.W.: Fast effective rule induction. In: International Conference on Machine Learning (ICML), pp. 115–123 (1995)

Ciaperoni, M., Xiao, H., Gionis, A.: Efficient exploration of the rashomon set of rule-set models. In: Proceedings Of The 30th ACM SIGKDD Conference On Knowledge Discovery And Data Mining, pp. 478–489 (2024)

Dash, S., Günlük, O., Wei, D.: Boolean decision rules via column generation. In: Advances in Neural Information Processing Systems (NeurIPS), vol. 31 (2018)

Ghosh, B., Meel, K.S.: IMLI: An incremental framework for MaxSAT-based learning of interpretable classification rules. In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, pp. 203–210 (2019)

Ghosh, B., Malioutov, D., Meel, K.S.: Efficient learning of interpretable classification rules. Journal of Artificial Intelligence Research **74**, 1823–1863 (2022)

Hühn, J., Hüllermeier, E.: FURIA: an algorithm for unordered fuzzy rule induction. Data Mining and Knowledge Discovery **19**(3), 293–319 (2009)

Ignatiev, A., Lam, E., Stuckey, P.J., Marques-Silva, J.: A scalable two stage approach to computing optimal decision sets. In: AAAI Conference on Artificial Intelligence (2021)

Ignatiev, A., Pereira, F., Narodytska, N., Marques-Silva, J.: A SAT-based approach to learn explainable decision sets. In: International Joint Conference on Automated Reasoning, pp. 627–645 (2018). Springer

Janosi, A., Steinbrunn, W., Pfisterer, M., Detrano, R.: Heart Disease [Dataset]. UCI Machine Learning Repository (1988). https://doi.org/10.24432/C52P4X

Kohavi, R., Becker, B.: Adult [Dataset]. UCI Machine Learning Repository (1996). https://doi.org/10.24432/C5XW20

Lakkaraju, H., Bach, S.H., Leskovec, J.: Interpretable decision sets: A joint framework for description and prediction. In: Proceedings of the 22nd ACM KDD International Conference on Knowledge Discovery and Data Mining, pp. 1675–1684 (2016)

Lu, J., Ester, M.: An active approach for model interpretation. In: NeurIPS 2019 Workshop on Human-Centric Machine Learning (2019)

Malioutov, D., Meel, K.S.: MLIC: A MaxSAT-based framework for learning interpretable classification rules. In: International Conference on Principles and Practice of Constraint Programming, pp. 312–327 (2018). Springer

Mita, G., Papotti, P., Filippone, M., Michiardi, P.: LIBRE: Learning interpretable boolean rule ensembles. In: International Conference on Artificial Intelligence and Statistics (AISTATS), pp. 245–255 (2020)

National Institute of Justice: NIJ Recidivism Forecasting Challenge Dataset. https://nij.ojp.gov/funding/recidivism-forecasting-challenge (2021)

OpenML: Pima Indians Diabetes Database. https://www.openml.org/d/43483. Originally from the UCI Machine Learning Repository (2014)

OpenML: FICO-HELOC-cleaned (2026). https://www.openml.org/d/45023

ProPublica: COMPAS Recidivism Risk Score Data and Analysis. GitHub repository. Accessed 2026-01-08 (2016). https://github.com/propublica/compas-analysis

Wang, T.: Multi-value rule sets for interpretable classification with feature-efficient representations. In: Advances in Neural Information Processing Systems (NeurIPS), vol. 31 (2018)

Wang, T., Rudin, C., Doshi-Velez, F., Liu, Y., Klampfl, E., MacNeille, P.: A bayesian framework for learning rule sets for interpretable classification. The Journal of Machine Learning Research **18**(1), 2357–2393 (2017)

Wang, T., Rudin, C., Doshi-Velez, F., Liu, Y., Klampfl, E., MacNeille, P.: In-Vehicle Coupon Recommendation [Dataset] (2020). https://doi.org/10.24432/C5GS4P

Wang, Z., Zhang, W., Liu, N., Wang, J.: Scalable rule-based representation learning for interpretable classification. In: Advances in Neural Information Processing Systems (NeurIPS), vol. 34 (2021)

Yang, F., He, K., Yang, L., Du, H., Yang, J., Yang, B., Sun, L.: Learning interpretable decision rule sets: A submodular optimization approach. In: Advances in Neural Information Processing Systems (NeurIPS), vol. 34 (2021)

Yu, J., Ignatiev, A., Stuckey, P.J., Le Bodic, P.: Computing optimal decision sets with sat. In: International Conference on Principles and Practice of Constraint Programming, pp. 952–970 (2020). Springer

Yang, L., Leeuwen, M.: Probabilistic rule sets ready for interactive machine learning. In: AAAI-22 Workshop on Interactive Machine Learning (IML@AAAI'22) (2022)

Zhang, G., Gionis, A.: Diverse rule sets. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 1532–1541 (2020)