A IMPLEMENTATION DEFAULTS

Table 5 contains existing implementation defaults used in our experiments. They have been obtained from the current versions of the implementations. We analyze algorithms from the following algorithm implementations: Elastic Net: `glmnet` [18] , Decision Trees: `rpart` [46], Random Forest: `ranger` [56], SVM: `LibSVM` via `e1071` ([9], [35]) and `xgboost` [10]. We investigate `HNSW` [32] as an approximate k-Nearest-Neighbours algorithm. Additional details on the exact meaning of the different hyperparameters can be obtained from the respective software's documentation. We assume that small differences due to implementation details e.g. between the `LibSVM` and `sklearn` implementations exist, but try to compare to existing default settings nonetheless, as they might serve as relevant baselines.

B EVOLUTIONARY SEARCH

The main components of the evolutionary search are given in Section 4.2. This appendix only defines the mutation operations. We use a number of mutation operators, but not all mutations can be applied to all candidates. Given a parent, we first determine the mutations that will lead to valid offspring, after which we apply one chosen uniformly at random. The mutation operators are:
Node Insertion: Pick a node in the tree and add an operator node between that node and its parent node. If necessary, additional input to the new node is generated by randomly selecting terminals.
Shrink Node: Select a node and replace it with one of its subtrees.
Node Replacement: Replace a randomly chosen node by another node of the same type and arity.
Terminal Replacement: Replace a terminal with a different terminal of the same type (i.e. <I> or <F>).
Mutate Ephemeral: Change the value of an ephemeral constant (i.e. c_i or c_f) with Gaussian noise proportional to the ephemeral's value. For the integer ephemeral, the change is rounded and can not be zero.
None of the mutations that work on operators work on the <configuration> operator.
In order to define the search space for symbolic formulas, we define a grammar composed of terminal symbols and operators. Operators can take one or multiple terminals or operators as input and produce a single output. We consider two types of terminal symbols: ephemeral constants and meta-features. This allows for a very flexible description of the search space.

$\mu + \lambda$ vs. random search. Figure 7 depicts optimization traces of $\mu + \lambda$ and random search across 10 replications on all datasets. Shorter EA traces occur due to early stopping. Genetic programming seems to consistently yield better results.

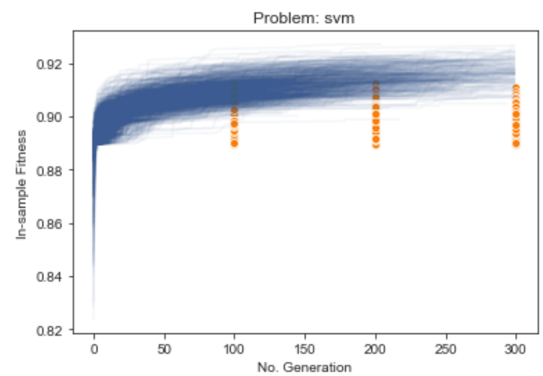


Figure 7: In-sample fitness scores of $\mu + \lambda$ (blue) and 100, 200 and 300 generations equivalent of random search (orange).

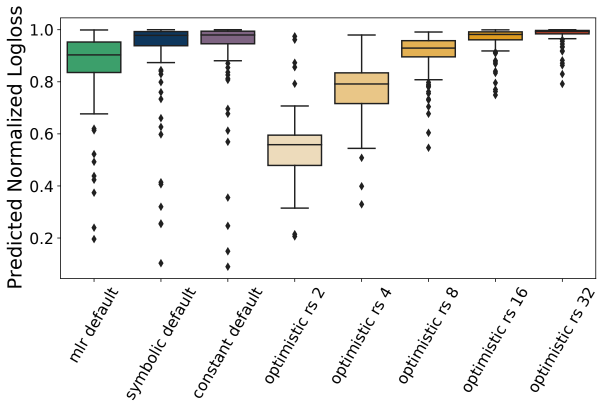
Algorithm	Default	
Elastic Net	glmnet:	$\alpha : 1, \lambda : 0.01$
Decision Tree	rpart:	$cp : 0.01, max.depth : 30, minbucket : 1, minsplit : 20$
Random Forest	ranger:	$mtry : \sqrt{po}, sample.fraction : 1, min.node.size : 1$
SVM	e1071: sklearn:	$C : 1, \gamma : \frac{1}{po}$ $C : 1, \gamma : \frac{1}{p{*}xvar}$
Approx. kNN	mlr:	$k : 10, M : 16, ef : 10, efc : 200$
Gradient Boosting	xgboost:	$\eta : 0.1, \lambda : 1, \gamma : 0, \alpha : 0, subsample : 1, max\_depth : 3, min\_child\_weight : 1, colsample\_bytree : 1, colsample\_bylevel : 1$

Table 5: Baseline b): Existing defaults for algorithm implementations. Fixed parameters described in Table 3 apply

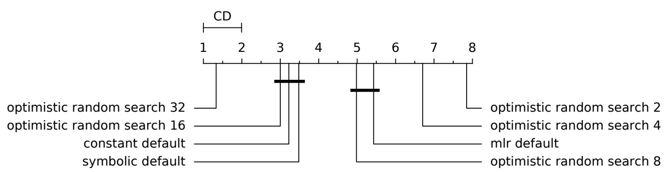
C EXPERIMENTAL RESULTS

The following section describes the results of the Experiments conducted to answer **RQ1** and **RQ2** across all other algorithms analyzed in this paper. Results and a more detailed analysis for the SVM can be obtained from section 6.2.

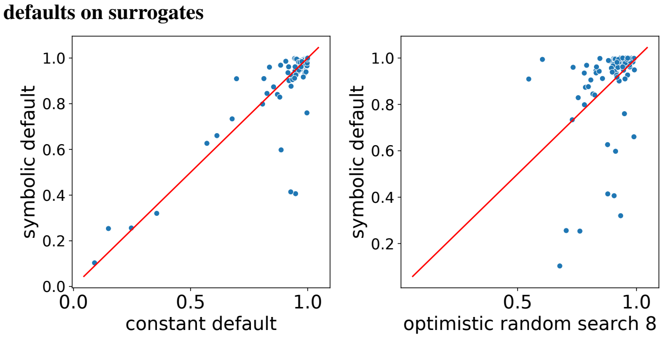
C.1 Elastic Net



(a) Symbolic, static and implementation defaults, comparing scaled logloss predicted by surrogates.



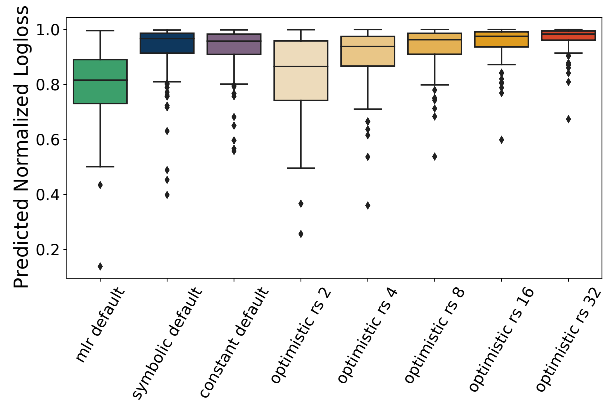
(b) Critical Differences Diagram of symbolic, static and implementation defaults on surrogates



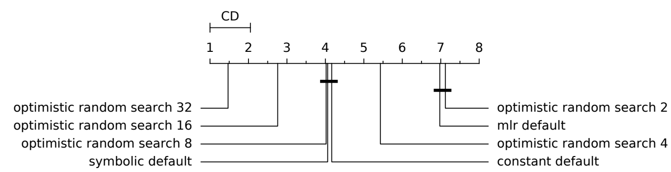
(c) Performance comparison of symbolic defaults to constant defaults (left) and budget 8 optimistic random search (right).

Figure 8: Results for the elastic net algorithm on surrogate data.

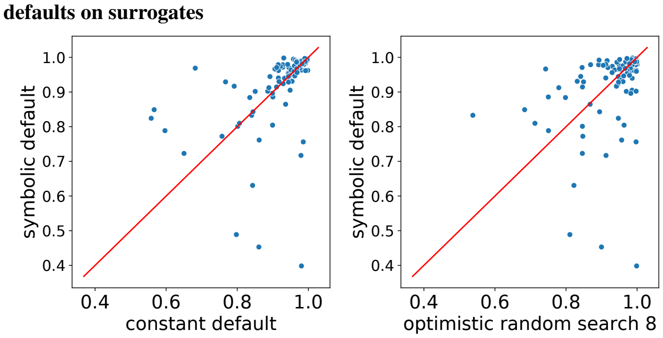
C.2 Decision Trees



(a) Symbolic, static and implementation defaults, comparing scaled logloss predicted by surrogates.



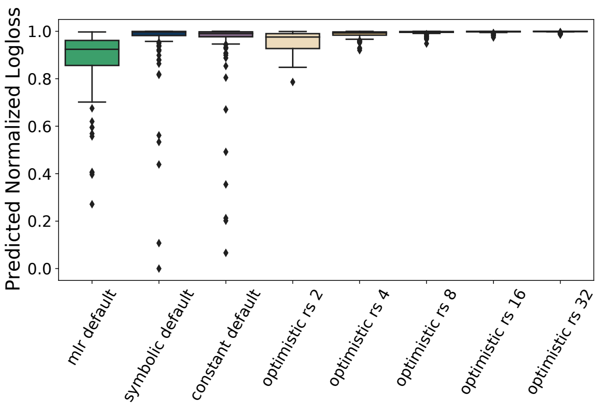
(b) Critical Differences Diagram of symbolic, static and implementation defaults on surrogates



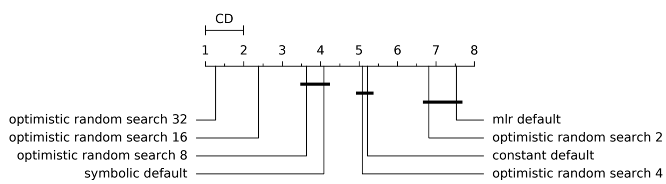
(c) Performance comparison of symbolic defaults to constant defaults (left) and budget 8 optimistic random search (right).

Figure 9: Results for the decision tree algorithm on surrogate data.

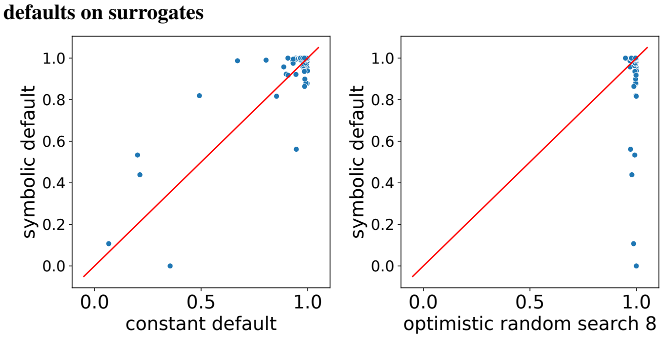
C.3 Approximate k-Nearest Neighbours



(a) Symbolic, static and implementation defaults, comparing scaled logloss predicted by surrogates.



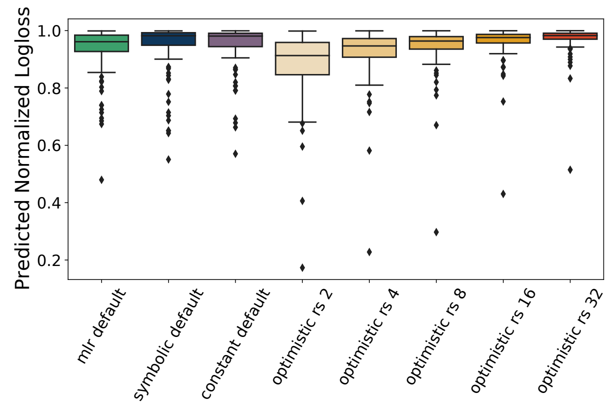
(b) Critical Differences Diagram of symbolic, static and implementation defaults on surrogates



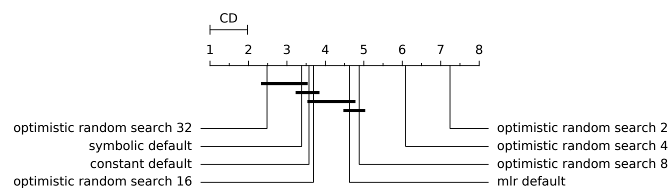
(c) Performance comparison of symbolic defaults to constant defaults (left) and budget 8 optimistic random search (right).

Figure 10: Results for the approximate k-nearest neighbours algorithm on surrogate data.

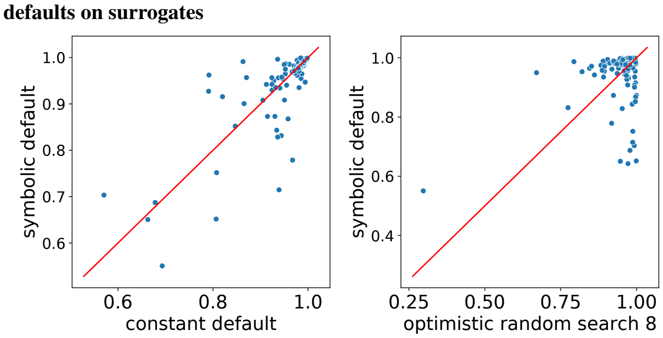
C.4 Random Forest



(a) Symbolic, static and implementation defaults, comparing scaled logloss predicted by surrogates.



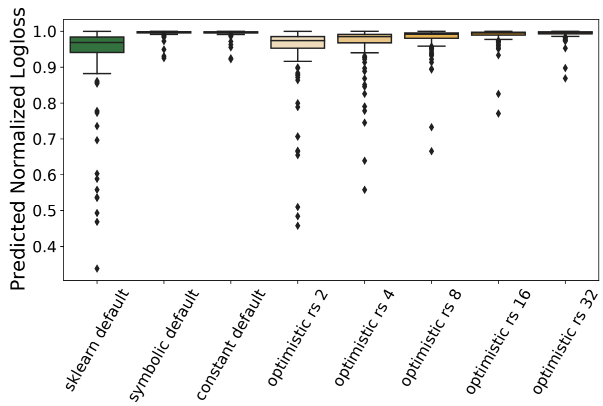
(b) Critical Differences Diagram of symbolic, static and implementation defaults on surrogates



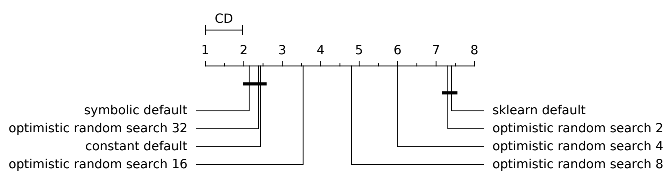
(c) Performance comparison of symbolic defaults to constant defaults (left) and budget 8 optimistic random search (right).

Figure 11: Results for the random forest algorithm on surrogate data.

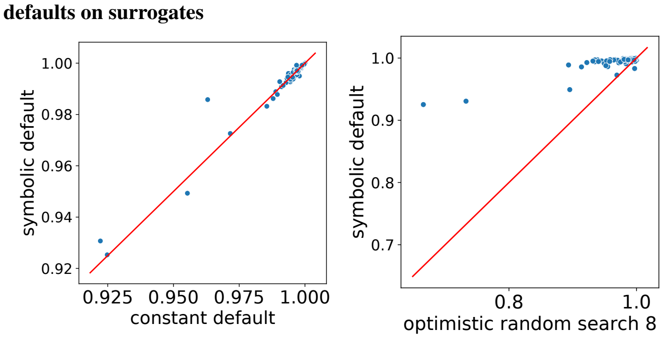
C.5 eXtreme Gradient Boosting (XGBoost)



(a) Symbolic, static and implementation defaults, comparing scaled logloss predicted by surrogates.



(b) Critical Differences Diagram of symbolic, static and implementation defaults on surrogates



(c) Performance comparison of symbolic defaults to constant defaults (left) and budget 8 optimistic random search (right).

Figure 12: Results for the XGBoost algorithm on surrogate data.

D REAL DATA EXPERIMENTS

In analogy to the presentation of the results for the SVM of the main text, we present results for Decision Tree and Elastic Net here.

D.1 Decision Tree

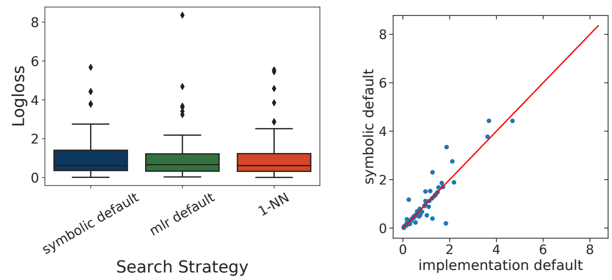


Figure 13: Results for the decision tree algorithm. Comparison of symbolic and implementation default using log-loss across all datasets performed on real data. Box plots (right) and scatter plot (left)

D.2 Elastic Net

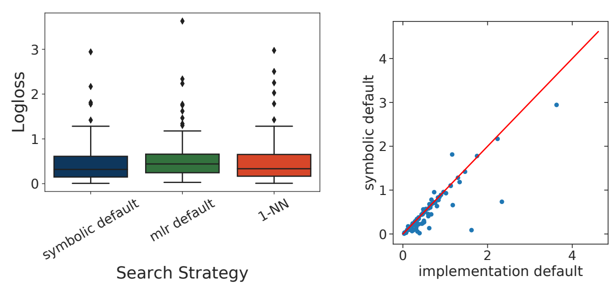


Figure 14: Results for the Elastic Net algorithm. Comparison of symbolic and implementation default using log-loss across all datasets performed on real data. Box plots (right) and scatter plot (left)