MEASE: Mixed Type Evolutionary Algorithm for Exceptional Survival Model Mining

Bruno Leal Fonseca, Déborah Brito Yamamoto, Renato Vimieiro
Departamento de Ciência da Computação
Universidade Federal de Minas Gerais
Belo Horizonte, Brazil
{bruno.leal.fonseca, dbyamamoto, rvimieiro}@dcc.ufmg.br

***Abstract*—Exceptional survival subgroup discovery aims to identify interpretable descriptions of subpopulations whose time-to-event behavior differs from a reference group under censoring. Most descriptive approaches for survival data still rely on user-defined discretization of numerical variables, which is costly, dataset-dependent, and can strongly affect the discovered patterns. This paper proposes an evolutionary exceptional model mining method for survival analysis that induces numerical constraints during the search, enabling mixed categorical and numerical descriptions without manual discretization. The Mixed Type Evolutionary Algorithm for Exceptional Survival Model Mining (MEASE) maintains a redundancy-aware top-K archive and uses adaptive crossover and mutation rates with restart mechanisms to balance exploration and exploitation while promoting diversity. Experiments on six real-world benchmark survival datasets compare the proposed approach against representative baselines methods. The results show that the proposed method achieves competitive exceptionality while typically improving subgroup coverage, without requiring any user preprocessing on the datasets.**

***Index Terms*—subgroup discovery, exceptional model mining, evolutionary algorithms, survival analysis**

I. Introduction

Survival analysis (SA) studies time-to-event data under censoring and is widely used in domains such as medicine and engineering. While a major research line in SA focuses on building predictive models to answer *if* and *when* an event will occur, many practical settings require *descriptive* insights: *Are there subpopulations with unusual survival behavior?* and *Which attribute conditions characterize such exceptional groups?*

Subgroup Discovery (SD) and Exceptional Model Mining (EMM) address these descriptive questions by searching for human-interpretable rule descriptions whose covered instances exhibit behavior that contrasts with a reference group (e.g., the population or the complement) [1], [2]. In the context of SA, this corresponds to discovering subgroups whose Kaplan-Meier survival curves are statistically different from a baseline, commonly assessed by the *log-rank test*. However, a persistent limitation of many SD/EMM approaches is their dependence on *discretized* descriptive attributes: numerical variables are typically discretized in a manual and iterative preprocessing step, which is costly, varies from dataset to dataset, and can strongly influence the discovered patterns. There are surveys that review several state-of-the art approaches for this problem and methods to deal with numerical attributes [3].

This work investigates the problem of mining *exceptional survival subgroups* in datasets with *mixed* (categorical and numerical) attributes *without user-defined discretization*.

This raises a question: Is it possible to design an SD/EMM approach for survival data that (i) supports censoring; (ii) discovers diverse and interpretable exceptional survival patterns; and (iii) handles *mixed* (categorical and numerical) attributes *without* requiring user-defined discretization?

The Mixed Type Evolutionary Algorithm for Exceptional Survival Model Mining (MEASE) is an evolutionary SD/EMM approach that (i) induces numerical constraints during the search, (ii) returns a Top-K set of non-redundant exceptional rules, and (iii) balances exceptionality, coverage, and diversity.

While several SD/EMM algorithms exist, EsmamDS [4] currently represents the state of the art in relation to SA applications. MEASE differs from EsmamDS in three fundamental aspects. First, MEASE operates directly on continuous numerical attributes by evolving interval-based descriptions, eliminating the need for manual discretization required by most SD/EMM algorithms; this step is dataset-dependent and known to influence the quality of the discovered patterns. Second, the two methods adopt fundamentally different search paradigms. EsmamDS employs a sequential covering strategy based on Ant Colony Optimization (ACO), in which a single subgroup is discovered per iteration and the search is restarted over the remaining uncovered instances. MEASE, in contrast, evolves an entire population of candidate rules simultaneously through genetic operators, allowing the search to explore multiple regions of the description space in parallel. Third, these paradigms entail different evaluation costs during search: EsmamDS must re-evaluate a partial description each time a condition is added or pruned within the ACO construction step, whereas in MEASE each individual is evaluated once per generation. The cost of evaluating a subgroup in MEASE is potentially lower because EsmamDS reevaluates a subgroup every time a new description is added or pruned; during the execution, a subgroup of extensive length will, naturally, be pruned to optimize interpretability. However, these successive reevaluations can be expensive, compared to MEASE that only evaluated a subgroup once, and then evaluates it's decedents in future generations.

II. Subgroup Discovery and Survival Analysis

Discovering exceptional patterns or subgroups helps to explain why a subset of samples exhibits behavior that differs from the rest of the population. Such findings point to relevant subsets for which alternative interventions or decision strategies may be more appropriate, potentially improving survival prospects and reducing the likelihood of the target event.

A. Survival Analysis

SA has been used to investigate, for example, the influence of genetic factors on cancer subtypes by identifying traits associated with better or worse survival outcomes [5]. In scenarios like this, SD and EMM are particularly useful: labeled classes enable the search for descriptive patterns that distinguish groups, while time-to-event data allow the mining of characteristics that make certain cases exceptional. Such insights can support research on targeted treatments for specific patient groups, which is valuable because adequately addressing particular cases can improve survival outcomes.

Table I shows a small subset of attributes and samples from a discretized dataset that will be used in the experimental section. Two pieces of information are essential for SA: the `event` indicator and the `survival-time` (time-to-event) variable. The `event` column indicates whether the target event was observed (`1`) or the sample was censored (`0`), typically because the study ended, the participant withdrew, or another competing event prevented further observation. The `survival-time` column indicates the duration of follow-up until the event occurred or censoring took place.

TABLE I
Subset of samples and attributes from one of the datasets used in the experiments

age	condition	site	n-staging	survival-time	event
44	condition-1	type-4	class-2	1307	0
48	condition-1	type-1	class-2	230	1
67	condition-1	type-2	class-1	763	1
58	condition-2	type-4	class-3	172	1
69	condition-1	type-1	class-2	1455	0
75	condition-1	type-4	class-0	1234	0

One of the most widely used estimators of the survival function is the Kaplan-Meier (KM) method [6].

To assess whether two KM curves differ, the *log-rank test* can be used, which signals exceptionality during the search.

B. Subgroup Discovery

In SD, a *rule* or subgroup *description* is typically defined as a conjunction of *selectors* (attribute-value constraints). Let I denote the set of all available selectors over the dataset attributes. Using Table I as an example, $I = \{\text{condition} = \text{condition-1}, \text{condition} = \text{condition-2}, \ldots, \text{n-staging} = \text{type-2}, \text{n-staging} = \text{type-3}\}$. A description d is any conjunction of one or more selectors from I. For instance, $d = \{\text{condition} = \text{condition-1}, \text{condition} = \text{condition-2}, \text{site} = \text{type-1}\} \equiv \text{condition} \in \{\text{condition-1}, \text{condition-2}\} \wedge \text{site} = \text{type-1}$

To mitigate limitations associated with greedy search heuristics, such as beam search, widely applied in rule learner SD algorithms, *SSDP* [7] was proposed, followed by *SSDP+* [8]. Instead of *beam search*, these methods employ evolutionary search, specifically, genetic algorithms with tournament selection (GATS) which returns the Top-K highest-quality rules. *Crossover* and *mutation* promote both intensification around promising descriptions and diversification toward novel patterns.

SSDP introduces adaptive *crossover* and *mutation* rates. When no improvements are observed in the Top-K set, *mutation* is increased and *crossover* is reduced to encourage exploration; when new high-quality rules are incorporated into Top-K, *crossover* is favored to intensify the search around promising regions. Although *SSDP+* substantially reduces redundancy among returned rules by grouping descriptions, it still assumes categorical attributes and does not directly handle numerical variables.

Overall, the methods discussed so far require a preprocessing step for numerical attributes, which can be labor-intensive and iterative because users often need multiple runs to obtain a satisfactory discretization. Note that, if all possible constraints for the attribute age in Table I were explored, an expressive complexity would be added in algorithm's search, because the continuous values would be handled as discretized, but with a large amount of possibilities. To address this limitation, an MA thesis introduces *EASD* [9], which performs discretization of numerical attributes *during* the search process. *EASD* targets labeled datasets and, similarly to *SSDP*/*SSDP+*, adopts an evolutionary strategy optimized with WRAcc. Numerical attributes are discretized by creating constraints (e.g., intervals) so that covered samples of a target class are better distinguished from the other individuals of the remaining classes.

C. Exceptional Model Mining on Survival Analysis

Motivated by the need for interpretable, rule-based approaches to SA, EMM and SD have been investigated as alternatives to conventional predictive models. In this context, the Exceptional Survival Model Ant Miner (*ESMAM*) was proposed as a distinctive contribution to the state of the art [10]. More recently, the authors refined the approach and introduced *EsmamDS* [4]. Both algorithms are based on ACO to mine survival-related patterns, aiming to discover attribute descriptions whose induced survival curves are exceptional when compared with the overall population or with their complements.

EsmamDS uses an Ant Colony Optimization strategy to construct rule descriptions, guided by a *log-rank* based quality measure (e.g., $1 - p\text{-}value$). A known issue is that very small subgroups may achieve extremely low $p\text{-}values$, motivating a coverage-weighted scoring to avoid overly specific patterns; selector-frequency penalties are also used to reduce redundancy. Still, EsmamDS operates on discrete descriptions and requires discretization when numerical attributes are present.

Algorithm 1 MEASE Pseudocode

Require: Dataset D, Top-K size K, max generations G_{max}, population size P, max stagnation generations G_{noImp}, max restarts $R_{\max}$, restart fraction ρ, minimum support $minSupp$, maximum support $maxSupp$, support importance α, redundancy threshold τ, tolerance ϵ
Ensure: Top-K set $\mathcal{T}$ of best non-redundant rules

$g \leftarrow 0$; $r \leftarrow 0$; $c \leftarrow 0$ ▷ c: stagnation counter
$(p_c, p_m) \leftarrow (0.6, 0.4)$ ▷ crossover/mutation rates
$\mathcal{T} \leftarrow \emptyset$; initialize heap/index structures
$\mathcal{P} \leftarrow initPopulation(D, P)$
while $g < G_{max}$ **and** $r < R_{\max}$ **do**
  $rulesFitness \leftarrow evaluate(\mathcal{P}, D, \alpha, minSupp, maxSupp)$
  $\mathcal{P} \leftarrow CrossoverMutation(\mathcal{P}, rulesFitness, D, p_c, p_m)$
  $rulesFitness \leftarrow evaluate(\mathcal{P}, D, \alpha)$
  $\mathcal{T} \leftarrow$ UPDATETOPK$(\mathcal{P}, rulesFitness, \mathcal{T}, K, \tau, \epsilon)$
  $(p_c, p_m) \leftarrow adaptCrossoverMutationRates(\mathcal{T}, p_c, p_m)$
  if not $checkImprovement(\mathcal{T})$ **and** $p_m = 1.0$ **then**
    $c \leftarrow c + 1$
  else
    $c \leftarrow 0$
  end if
  if $c \geq G_{noImp}$ **and** $p_m = 1.0$ **then**
    $\mathcal{P} \leftarrow restartPopulation(\mathcal{P}, rulesFitness, \rho, D)$
    $(p_c, p_m) \leftarrow (0.6, 0.4)$; reset stagnation bookkeeping
    $c \leftarrow 0$; $r \leftarrow r + 1$
  end if
  $g \leftarrow g + 1$
end while
return $\mathcal{T}$

Algorithm 2 UPDATETOPK$(\mathcal{P}, rulesFitness, \mathcal{T}, K, \tau, \epsilon)$

Require: Population $\mathcal{P}$, fitness vector $rulesFitness$, Top-K set $\mathcal{T}$, size K, redundancy threshold τ, tolerance ϵ
Ensure: Updated Top-K set $\mathcal{T}$

for $i \leftarrow 1$ to $|\mathcal{P}|$ **do**
  $rule \leftarrow \mathcal{P}[i]$; $fit \leftarrow rulesFitness[i]$
  $k \leftarrow$ GETATTRIBUTESKEY$(rule)$
  if $k \in \mathcal{T}$ **then** ▷ Update existing rule with better intervals
    if $fit > \mathcal{T}[k].score + \epsilon$ **then**
      $\mathcal{T}[k] \leftarrow (fit, rule)$
      Update heap: remove $(oldFit, k)$, insert (fit, k)
    end if
    continue
  end if
  if REDUNDANCYTEST$(rule, fit, \mathcal{T}, \tau, \epsilon)$ **then**
    continue
  end if
  if $|\mathcal{T}| < K$ **or** $fit >$ GETWORST$(\mathcal{T}).score + \epsilon$ **then**
    $\mathcal{T}[k] \leftarrow (fit, rule)$
    HEAPPUSH(fit, k)
    while $|\mathcal{T}| > K$ **do**
      PRUNESTALEHEAPENTRIES()
      $(worstFit, worstKey) \leftarrow$ HEAPPOP()
      Remove $worstKey$ from $\mathcal{T}$
    end while
  end if
end for
if heap size $> |\mathcal{T}|$ **then** ▷ Cleanup if too many stale entries
  REBUILDHEAP$(\mathcal{T})$
end if
return $\mathcal{T}$

Despite these advances, both *ESMAM* and *EsmamDS* share a limitation with several descriptive SD/EMM approaches discussed earlier: they operate on discrete or categorical attributes and therefore require discretization when numerical variables are present.

III. METHODOLOGY

MEASE is an evolutionary SD/EMM approach for mining *exceptional survival subgroups* from datasets containing mixed categorical and numerical attributes. The method is inspired by the evolutionary search strategy of SSDP+ [8] returning the Top-K best rules of all generations and by the idea of inducing numerical constraints during search as in EASD [9], while targeting the exceptional-survival setting of EsmamDS [4].

Each individual represents a rule (subgroup description) for both categorical and numerical attributes. Given a candidate rule, Kaplan-Meier models are fitted for the covered subgroup and a baseline group (complement or population) to quantify exceptionality by a log-rank test $p\text{-}value$ score. A redundancy-aware Top-K archive is maintained across generations to retain high-quality yet diverse rules according to a similarity threshold.

Algorithm 1 is a resumed representation of the proposed methodology; Algorithm 2 is a detailed description of how the Top-K best rules are updated in every generation. MEASE was implemented in Python and the public GitHub repository is available in: https://github.com/debyyamamoto/MEASE

At each generation, all individuals are evaluated, the Top-K archive is updated, and the next population is generated through *crossover* and *mutation*. The rates are initialized as $p_c = 0.60$ and $p_m = 0.40$ and are adaptively adjusted to balance exploitation and exploration. When no improvement is observed in the Top-K archive, exploration is encouraged by increasing p_m and decreasing p_c by 0.20; conversely, when improvements occur, p_c is increased and p_m is reduced by the same step.

To mitigate premature convergence, a restart strategy was employed, triggered after G_{noImp} consecutive generations without Top-K improvement. At restart, an elitist fraction (10%) of the current population is preserved and the remaining individuals are re-initialized at random. Restarts are performed at most $R_{\max}$ times; the search terminates when $R_{\max}$ is reached or the maximum number of generations $G_{\max}$ is met, returning the final Top-K archive.

During the update step in Algorithm 2, each candidate rule $rule \in \mathcal{P}$ is compared against the current archive. If the archive already contains a rule with the same *attribute set* as a $rule$ (i.e., the same description schema), the best-scoring constraints/categories are kept by replacing the archived rule when the new fitness improves it by at least ϵ.

If the attribute set is new and either the archive is not full or the candidate outperforms the current worst entry (by at least ϵ), the candidate is inserted and the worst rule is discarded to maintain a fixed-size Top-K archive.

Adaptive rates and restarts do not guarantee diversity: high-

TABLE II
CHARACTERISTICS OF 6 SURVIVAL DATASETS USED IN THE EXPERIMENTAL STUDY.

dataset	$\|n\|$	$\|A\|$	$\%cens$	Subject of research	Event
breast-cancer	196	80	73.98	Node-Negative breast cancer	distant metastasis
cancer	168	7	27.98	Advanced lung cancer	death
carcinoma	193	8	27.46	Carcinoma of the oropharynx	death
lung	901	8	37.40	Early lung cancer	death
mgus2	1338	7	29.90	Monoclonal gammopathy	death
veteran	137	6	6.57	Lung cancer	death

scoring regions may still yield many near-duplicate rules. Therefore, redundancy control was enforced when updating the archive. For each candidate, coverage similarity is measured against archived rules using the Jaccard coefficient between their covered-instance sets. If similarity exceeds a threshold τ, the candidate is treated as redundant and is not inserted (unless it improves the corresponding archived rule by at least ϵ).

IV. EXPERIMENTS

Experiments were made using a subset of the SA datasets used in EsmamDS [4]. Some of these consisted of categorical datasets, others had both numerical and categorical attributes. The Table II shows the datasets used in the experiments. The columns n indicates the number of samples; A the number of attributes, disregarding survival time and event occurrence; $\%cens$ is the percentage of samples that were censored in the study.

Experiments were conducted on an AMD Ryzen 5 4600H with Radeon Graphics (3.00 GHz) 6 cores 12-threads; 32,0 GB.

The exceptionality calculation took as reference the rule's *complement* to estimate the fitness through *log-rank test*. For the purpose of getting rules that are not too exceptional, a support threshold was established to enhance the rule's support importance parameter α. The minimum and maximum support were established at *0.05* and *0.55*, respectively. Therefore, rules that cover too many or not enough samples are eliminated from Top-K. It was observed that without this threshold the best rules tended to converge to the population survival curve whereas its complement was just a portion small portion of the dataset. The opposite is observed with rules with low coverage and complements that tended to the population survival curve.

MEASE was evaluated over 30 executions per dataset, alternating seeds and recording mean runtime, mean RAM, and peak RAM consumption. Computational performance results are reported in Table III. When all executions were performed, statistics about metrics intrinsically related to the rule's exceptionality were estimated. Our goal is to compare our approach with the state of the art algorithms. Table IV shows the parameters used in each algorithm. In addition, a *Wilcoxon statistical test* was performed on the subgroup quality metrics to evaluate if there is a statistically significant difference between the rules returned by MEASE and the others. Table V presents the results of the paired *Wilcoxon statistical test*, including the mean and standard deviation for each metric. In the MEASE rows, there is a report of both the metric values and the statistical test outcomes, indicating whether our approach significantly outperformed, matched, or underperformed the baselines.

TABLE III
PERFORMANCE RESULTS

dataset	mean-runtime	mean-ram-mb	peak-ram-mb
breast-cancer	234.31 ± 47.12	212.78 ± 0.75	212.8 ± 0.74
cancer	240.89 ± 67.6	207.53 ± 0.23	207.55 ± 0.21
carcinoma	209.86 ± 65.21	207.56 ± 0.2	207.58 ± 0.19
lung	440.2± ± 12.32	210.5 ± 0.36	210.52 ± 0.34
mgus2	408.96 ± 101.09	211.92 ± 0.57	211.96 ± 0.47
veteran	199.1 ± 50.94	206.94 ± 0.27	206.96 ± 0.26

The quality of the discovered subgroups is assessed using a set of metrics that capture different aspects of performance: (i) *exceptionality* (ε) measures the proportion of subgroups whose survival distribution is statistically different from the baseline, and is computed as the fraction of rules for which the *log-rank* test between the subgroup and the baseline yields a p-value less than or equal to a significance level of *0.05*. (ii) Interpretability is evaluated through *length*, defined as the average number of conditions (attributes) in a subgroup description. (iii) Generality is measured by *sgCov*, which represents the average relative coverage of individual subgroups, computed as $|SG_i|/|D|$; *stCov* is defined as $|\bigcup SG|/|D|$, indicating the proportion of instances covered by at least one subgroup. (iv) Redundancy is quantified through four complementary measures: *description redundancy* (ρ_D), computed as the average pairwise similarity between subgroup descriptions using $|I_i \cap I_j| / \min(|I_i|, |I_j|)$; *coverage redundancy* (ρ_C), defined analogously using the sets of covered instances; *cover redundancy* (CR), which captures instance-level redundancy by averaging the normalized excess number of times an instance is covered by multiple subgroups; and *model redundancy* (ρ_M), defined as the proportion of subgroup pairs whose survival models are not statistically distinguishable, i.e., pairs for which a *log-rank* test between their exclusive covered instances yields a p-value greater than a significance level of *0.05*.

A summary of the results obtained in our experiments is compiled in Table II, while some examples are depicted in Fig. 1[1]. Although the baseline used in these experiments was the complement, the population curve was added to the figure as a reference for all Top-K rules.

To investigate which dataset characteristics influence performance, Pearson (r) and Spearman (ρ) correlation coefficients were calculated between the number of instances and attributes and both mean runtime and its CV. A scatter plot was first inspected to assess the nature of each relationship. A monotonic trend between the number of instances and mean runtime

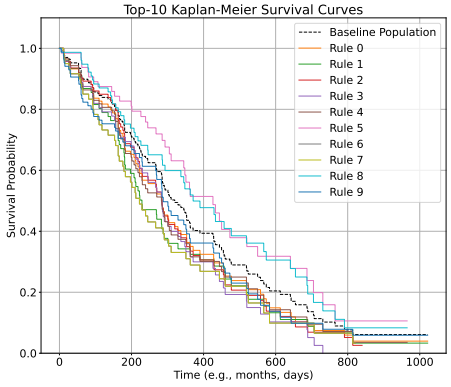
[1]The remainder of the results can be seen in our supplementary repository. https://github.com/debyyamamoto/MEASE

TABLE IV
INFORMATION ON THE ALGORITHMS COMPARED IN THIS WORK.

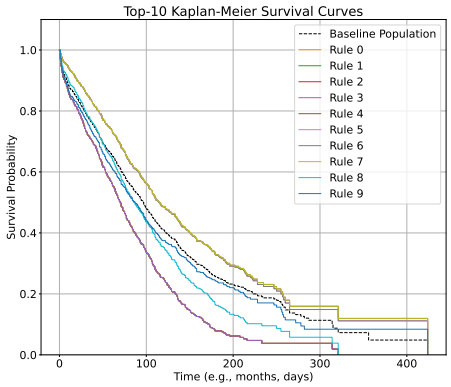
Algorithm	Search Strategy	Target Concept	Source [Reference]	Parameters (Value)
MEASE	GATS	KM moddel	MEASE Repository	$generations(500)$,$restart\text{-}gen(5)$, $restart\text{-}pop(5)$, $supportImportance$: $\alpha(0.10)$, $minSupp(0.05)$, $maxSupp(0.55)$, $ksize(10)$, $redundancy\text{-}threshold$: $\tau(0.90)$
EsmamDS-cpm	ACO	KM model	EsmamDS repository [4]	$\alpha(0.05)$, $nAnts(100)$, $minCov(0.05)$, $nConverg(5)$, $maxStag(40)$, $W(0.9)$, $L(10)$
Esmam-cpm	ACO	KM model	Esmam repository [10]	$\alpha(0.05)$, $nAnts(100)$, $minCov(0.05)$, $nConverg(5)$, $maxStag(40)$
BS-EMM-cpm	Beam Search	KM model	PySubgroup package [11]	$minCov(0.05)$, bs, $maxDepth$
BS-SD-cpm	Beam Search	Survival time	PySubgroup package [11]	$minCov(0.05)$, bs, $maxDepth$
LR-Rules	Sequential covering	KM model	LR-Rules repository [12]	–

TABLE V
EVALUATION METRICS PER DATASET: MEAN ± STD. SUPERSCRIPTS ON MEASE INDICATE WILCOXON SIGNED-RANK TEST OUTCOME ($\alpha = 0.05$) AGAINST EACH BASELINE IN COLUMN ORDER: ↑ WINS, ↓ LOSES, ○ DRAWS.

Dataset	**Algorithm**	ε	length	sgCov	stCov	ρ_D	ρ_C	CR	ρ_M
superscript order: E_{DS},E,BS_{EMM},BS_{SD},LR									
breast cancer	MEASE	$1.00_{0.00}^{\circ\circ\circ\circ\uparrow}$	$2.01_{\pm 0.23}^{\downarrow\downarrow\uparrow\uparrow\uparrow}$	$0.55_{\pm 0.00}^{\uparrow\uparrow\uparrow\uparrow\uparrow}$	$0.94_{\pm 0.07}^{\downarrow\downarrow\uparrow\uparrow\downarrow}$	$0.04_{\pm 0.04}^{\downarrow\circ\uparrow\uparrow\uparrow}$	$0.69_{\pm 0.08}^{\downarrow\downarrow\uparrow\uparrow\downarrow}$	$0.51_{\pm 0.01}^{\downarrow\downarrow\uparrow\uparrow\downarrow}$	$0.64_{\pm 0.20}^{\downarrow\circ\uparrow\uparrow\downarrow}$
	EsmamDS-cpm	$1.00_{\pm 0.00}$	$1.66_{\pm 0.33}$	$0.31_{\pm 0.08}$	$0.98_{\pm 0.02}$	$0.00_{\pm 0.01}$	$0.37_{\pm 0.11}$	$0.33_{\pm 0.06}$	$0.40_{\pm 0.10}$
	Esmam-cpm	$1.00_{\pm 0.00}$	$1.37_{\pm 0.15}$	$0.33_{\pm 0.04}$	$0.99_{\pm 0.02}$	$0.05_{\pm 0.03}$	$0.48_{\pm 0.04}$	$0.38_{\pm 0.05}$	$0.56_{\pm 0.08}$
	BS-EMM-cpm	$1.00_{\pm 0.00}$	$5.67_{\pm 0.00}$	$0.05_{\pm 0.00}$	$0.06_{\pm 0.00}$	$0.86_{\pm 0.00}$	$0.97_{\pm 0.00}$	$0.94_{\pm 0.00}$	$1.00_{\pm 0.00}$
	BS-SD-cpm	$1.00_{\pm 0.00}$	$5.78_{\pm 0.00}$	$0.08_{\pm 0.00}$	$0.13_{\pm 0.00}$	$0.70_{\pm 0.00}$	$0.79_{\pm 0.00}$	$0.87_{\pm 0.00}$	$1.00_{\pm 0.00}$
	LR-Rules	$0.81_{\pm 0.00}$	$4.44_{\pm 0.00}$	$0.13_{\pm 0.00}$	$1.00_{\pm 0.00}$	$0.10_{\pm 0.00}$	$0.19_{\pm 0.00}$	$0.41_{\pm 0.00}$	$0.39_{\pm 0.00}$
cancer	MEASE	$0.83_{\pm 0.11}^{\downarrow\downarrow\downarrow\circ\downarrow}$	$1.90_{\pm 0.21}^{\downarrow\downarrow\circ\circ\circ}$	$0.49_{\pm 0.04}^{\uparrow\uparrow\uparrow\uparrow\uparrow}$	$0.98_{\pm 0.02}^{\circ\uparrow\uparrow\uparrow\downarrow}$	$0.10_{\pm 0.05}^{\circ\uparrow\uparrow\uparrow\uparrow}$	$0.63_{\pm 0.08}^{\downarrow\uparrow\uparrow\downarrow\downarrow}$	$0.44_{\pm 0.04}^{\downarrow\uparrow\uparrow\uparrow\uparrow}$	$0.65_{\pm 0.16}^{\downarrow\circ\circ\downarrow\downarrow}$
	EsmamDS-cpm	$1.00_{\pm 0.00}$	$1.25_{\pm 0.19}$	$0.31_{\pm 0.03}$	$0.99_{\pm 0.02}$	$0.11_{\pm 0.10}$	$0.41_{\pm 0.06}$	$0.39_{\pm 0.05}$	$0.29_{\pm 0.11}$
	Esmam-cpm	$1.00_{\pm 0.00}$	$1.73_{\pm 0.10}$	$0.15_{\pm 0.03}$	$0.37_{\pm 0.16}$	$0.35_{\pm 0.05}$	$0.67_{\pm 0.02}$	$0.67_{\pm 0.07}$	$0.61_{\pm 0.12}$
	BS-EMM-cpm	$1.00_{\pm 0.00}$	$1.83_{\pm 0.00}$	$0.14_{\pm 0.00}$	$0.30_{\pm 0.00}$	$0.43_{\pm 0.00}$	$0.69_{\pm 0.00}$	$0.70_{\pm 0.00}$	$0.67_{\pm 0.00}$
	BS-SD-cpm	$0.83_{\pm 0.00}$	$1.83_{\pm 0.00}$	$0.13_{\pm 0.00}$	$0.36_{\pm 0.00}$	$0.30_{\pm 0.00}$	$0.48_{\pm 0.00}$	$0.64_{\pm 0.00}$	$0.40_{\pm 0.00}$
	LR-Rules	$1.00_{\pm 0.00}$	$1.82_{\pm 0.00}$	$0.20_{\pm 0.00}$	$1.00_{\pm 0.00}$	$0.27_{\pm 0.00}$	$0.37_{\pm 0.00}$	$0.49_{\pm 0.00}$	$0.33_{\pm 0.00}$
carcinoma	MEASE	$1.00_{\pm 0.00}^{\circ\circ\circ\circ\circ}$	$2.34_{\pm 0.13}^{\downarrow\downarrow\downarrow\downarrow\downarrow}$	$0.54_{\pm 0.01}^{\uparrow\uparrow\uparrow\uparrow\uparrow}$	$0.89_{\pm 0.07}^{\downarrow\circ\downarrow\downarrow\downarrow}$	$0.30_{\pm 0.10}^{\downarrow\circ\uparrow\uparrow\downarrow}$	$0.73_{\pm 0.07}^{\downarrow\circ\downarrow\downarrow\downarrow}$	$0.50_{\pm 0.01}^{\downarrow\circ\downarrow\downarrow\downarrow}$	$0.88_{\pm 0.14}^{\downarrow\circ\downarrow\downarrow\downarrow}$
	EsmamDS-cpm	$1.00_{\pm 0.00}$	$1.14_{\pm 0.20}$	$0.40_{\pm 0.06}$	$0.98_{\pm 0.02}$	$0.00_{\pm 0.00}$	$0.28_{\pm 0.19}$	$0.22_{\pm 0.13}$	$0.24_{\pm 0.18}$
	Esmam-cpm	$1.00_{\pm 0.00}$	$1.34_{\pm 0.37}$	$0.32_{\pm 0.10}$	$0.74_{\pm 0.30}$	–	–	$0.41_{\pm 0.21}$	–
	BS-EMM-cpm	$1.00_{\pm 0.00}$	$1.80_{\pm 0.00}$	$0.27_{\pm 0.00}$	$0.96_{\pm 0.00}$	$0.50_{\pm 0.00}$	$0.54_{\pm 0.00}$	$0.41_{\pm 0.00}$	$0.60_{\pm 0.00}$
	BS-SD-cpm	$1.00_{\pm 0.00}$	$1.60_{\pm 0.00}$	$0.32_{\pm 0.00}$	$0.96_{\pm 0.00}$	$0.35_{\pm 0.00}$	$0.34_{\pm 0.00}$	$0.34_{\pm 0.00}$	$0.40_{\pm 0.00}$
	LR-Rules	$1.00_{\pm 0.00}$	$1.00_{\pm 0.00}$	$0.33_{\pm 0.00}$	$0.99_{\pm 0.00}$	$0.00_{\pm 0.00}$	$0.00_{\pm 0.00}$	$0.01_{\pm 0.00}$	$0.33_{\pm 0.00}$
lung	MEASE	$1.00_{\pm 0.00}^{\circ\circ\circ\circ\circ}$	$2.67_{\pm 0.32}^{\downarrow\downarrow\downarrow\downarrow\downarrow}$	$0.55_{\pm 0.00}^{\uparrow\uparrow\uparrow\uparrow\uparrow}$	$0.95_{\pm 0.07}^{\downarrow\uparrow\circ\downarrow\downarrow}$	$0.13_{\pm 0.08}^{\downarrow\downarrow\downarrow\circ\downarrow}$	$0.69_{\pm 0.07}^{\downarrow\downarrow\downarrow\downarrow\downarrow}$	$0.50_{\pm 0.01}^{\downarrow\downarrow\downarrow\downarrow\downarrow}$	$0.23_{\pm 0.12}^{\uparrow\uparrow\circ\circ\uparrow}$
	EsmamDS-cpm	$1.00_{\pm 0.00}$	$1.12_{\pm 0.10}$	$0.39_{\pm 0.06}$	$1.00_{\pm 0.00}$	$0.00_{\pm 0.00}$	$0.43_{\pm 0.08}$	$0.32_{\pm 0.08}$	$0.36_{\pm 0.13}$
	Esmam-cpm	$1.00_{\pm 0.00}$	$1.00_{\pm 0.00}$	$0.33_{\pm 0.03}$	$0.84_{\pm 0.18}$	$0.00_{\pm 0.00}$	$0.57_{\pm 0.09}$	$0.44_{\pm 0.07}$	$0.62_{\pm 0.18}$
	BS-EMM-cpm	$1.00_{\pm 0.00}$	$1.00_{\pm 0.00}$	$0.35_{\pm 0.00}$	$0.94_{\pm 0.00}$	$0.00_{\pm 0.00}$	$0.39_{\pm 0.00}$	$0.30_{\pm 0.00}$	$0.27_{\pm 0.00}$
	BS-SD-cpm	$1.00_{\pm 0.00}$	$1.33_{\pm 0.00}$	$0.42_{\pm 0.00}$	$1.00_{\pm 0.00}$	$0.10_{\pm 0.00}$	$0.45_{\pm 0.00}$	$0.33_{\pm 0.00}$	$0.27_{\pm 0.00}$
	LR-Rules	$1.00_{\pm 0.00}$	$1.14_{\pm 0.00}$	$0.35_{\pm 0.00}$	$1.00_{\pm 0.00}$	$0.00_{\pm 0.00}$	$0.35_{\pm 0.00}$	$0.34_{\pm 0.00}$	$0.29_{\pm 0.00}$
mgus2	MEASE	$1.00_{\pm 0.00}^{\circ\circ\circ\circ\uparrow}$	$1.87_{\pm 0.12}^{\downarrow\downarrow\downarrow\downarrow\uparrow}$	$0.57_{\pm 0.01}^{\uparrow\uparrow\uparrow\uparrow\uparrow}$	$0.96_{\pm 0.03}^{\circ\uparrow\uparrow\uparrow\downarrow}$	$0.17_{\pm 0.06}^{\downarrow\downarrow\downarrow\downarrow\downarrow}$	$0.69_{\pm 0.03}^{\downarrow\downarrow\downarrow\downarrow\downarrow}$	$0.52_{\pm 0.01}^{\downarrow\downarrow\downarrow\downarrow\downarrow}$	$0.25_{\pm 0.07}^{\downarrow\downarrow\downarrow\downarrow\uparrow}$
	EsmamDS-cpm	$1.00_{\pm 0.00}$	$1.10_{\pm 0.11}$	$0.26_{\pm 0.04}$	$0.97_{\pm 0.02}$	$0.00_{\pm 0.00}$	$0.23_{\pm 0.10}$	$0.26_{\pm 0.08}$	$0.02_{\pm 0.03}$
	Esmam-cpm	$1.00_{\pm 0.00}$	$1.03_{\pm 0.06}$	$0.19_{\pm 0.00}$	$0.86_{\pm 0.06}$	$0.01_{\pm 0.01}$	$0.15_{\pm 0.02}$	$0.34_{\pm 0.04}$	$0.05_{\pm 0.02}$
	BS-EMM-cpm	$1.00_{\pm 0.00}$	$1.14_{\pm 0.00}$	$0.17_{\pm 0.00}$	$0.76_{\pm 0.00}$	$0.05_{\pm 0.00}$	$0.17_{\pm 0.00}$	$0.35_{\pm 0.00}$	$0.10_{\pm 0.00}$
	BS-SD-cpm	$1.00_{\pm 0.00}$	$1.29_{\pm 0.00}$	$0.17_{\pm 0.00}$	$0.74_{\pm 0.00}$	$0.10_{\pm 0.00}$	$0.23_{\pm 0.00}$	$0.40_{\pm 0.00}$	$0.05_{\pm 0.00}$
	LR-Rules	$0.79_{\pm 0.00}$	$2.00_{\pm 0.00}$	$0.16_{\pm 0.00}$	$1.00_{\pm 0.00}$	$0.07_{\pm 0.00}$	$0.15_{\pm 0.00}$	$0.32_{\pm 0.00}$	$0.29_{\pm 0.00}$
veteran	MEASE	$0.97_{\pm 0.06}^{\downarrow\downarrow\downarrow\downarrow\uparrow}$	$1.99_{\pm 0.10}^{\downarrow\downarrow\circ\uparrow\downarrow}$	$0.55_{\pm 0.02}^{\uparrow\uparrow\uparrow\uparrow\uparrow}$	$0.97_{\pm 0.05}^{\uparrow\uparrow\uparrow\uparrow\downarrow}$	$0.24_{\pm 0.05}^{\downarrow\downarrow\uparrow\uparrow\downarrow}$	$0.67_{\pm 0.06}^{\downarrow\downarrow\downarrow\uparrow\downarrow}$	$0.50_{\pm 0.02}^{\downarrow\circ\uparrow\uparrow\downarrow}$	$0.51_{\pm 0.17}^{\downarrow\circ\uparrow\uparrow\downarrow}$
	EsmamDS-cpm	$1.00_{\pm 0.00}$	$1.13_{\pm 0.11}$	$0.26_{\pm 0.02}$	$0.90_{\pm 0.01}$	$0.03_{\pm 0.03}$	$0.27_{\pm 0.01}$	$0.35_{\pm 0.03}$	$0.24_{\pm 0.05}$
	Esmam-cpm	$1.00_{\pm 0.00}$	$1.37_{\pm 0.15}$	$0.18_{\pm 0.03}$	$0.55_{\pm 0.15}$	$0.16_{\pm 0.09}$	$0.27_{\pm 0.11}$	$0.47_{\pm 0.13}$	$0.55_{\pm 0.23}$
	BS-EMM-cpm	$1.00_{\pm 0.00}$	$2.00_{\pm 0.00}$	$0.10_{\pm 0.00}$	$0.23_{\pm 0.00}$	$0.53_{\pm 0.00}$	$0.56_{\pm 0.00}$	$0.77_{\pm 0.00}$	$1.00_{\pm 0.00}$
	BS-SD-cpm	$1.00_{\pm 0.00}$	$2.50_{\pm 0.00}$	$0.08_{\pm 0.00}$	$0.21_{\pm 0.00}$	$0.70_{\pm 0.00}$	$0.69_{\pm 0.00}$	$0.79_{\pm 0.00}$	$1.00_{\pm 0.00}$
	LR-Rules	$0.82_{\pm 0.00}$	$1.36_{\pm 0.00}$	$0.19_{\pm 0.00}$	$1.00_{\pm 0.00}$	$0.09_{\pm 0.00}$	$0.22_{\pm 0.00}$	$0.32_{\pm 0.00}$	$0.35_{\pm 0.00}$



(a) `cancer`



(b) `mgus2`

Fig. 1. Kaplan–Meier survival curves produced by *MEASE*. Subfigures Subfigures indicate the Top-10 best discovered rules on two datasets: *cancer* and *mgus2*.

is visible, though not strictly linear, the values $r = 0.925$ and $\rho = 0.771$ diverge, suggesting that the largest datasets exert high leverage on the linear estimate. No consistent trend was observed between the number of attributes and runtime, indicating that runtime is governed by the number of instances evaluated per generation rather than the description space size.

Regarding memory consumption, MEASE was implemented in Python, whose runtime and standard scientific libraries (numpy, pandas, statsmodels) account for a static memory footprint of approximately 207MB. The algorithmic overhead above this baseline is at most 6MB across all datasets evaluated, with a standard deviation never exceeding 1MB, indicating stable and predictable memory usage regardless of dataset characteristics.

V. Conclusion

MEASE, an evolutionary SD/EMM approach for mining exceptional survival subgroups from datasets with mixed (categorical and numerical) attributes, avoiding user-defined discretization by inducing numerical constraints during the search. Across six benchmarks and 30 runs per dataset, MEASE achieved competitive exceptionality and coverage compared to the state of the art algorithms. At the same time, it was also able explicitly control redundancy, returning an informative list of patterns.

The statistical comparison (Wilcoxon signed-rank) indicates that MEASE is often comparable to EsmamDS on exceptionality-related metrics, although EsmamDS tends to yield lower description redundancy in several cases. Fig 1 makes this argument evident, the discovered subgroups have survival curves that differ from the baseline population, but in some datasets there is a clear model redundancy between Top-K rules. Future work includes integrating the baseline choice (population vs. complement) directly into the optimization objective and improving constraint induction to further increase diversity without sacrificing exceptionality. The search mechanism will be further improved, to avoid sensitivity to overly small or large subgroups.

References

[1] S. D. Bay and M. J. Pazzani, "Detecting group differences: Mining contrast sets," *Data mining and knowledge discovery*, vol. 5, no. 3, pp. 213–246, 2001.

[2] W. Duivesteijn, A. J. Feelders, and A. Knobbe, "Exceptional model mining: Supervised descriptive local pattern mining with complex target concepts," *Data Mining and Knowledge Discovery*, vol. 30, no. 1, pp. 47–98, 2016.

[3] M. Meeng and A. Knobbe, "For real: a thorough look at numeric attributes in subgroup discovery," *Data Mining and Knowledge Discovery*, vol. 35, no. 1, pp. 158–212, 2021.

[4] R. Vimieiro, J. B. Mattos, and P. S. de Mattos Neto, "Esmamds: A more diverse exceptional survival model mining approach," *Information Sciences*, vol. 690, p. 121549, 2025.

[5] H. H. Milioli, I. Tishchenko, C. Riveros, R. Berretta, and P. Moscato, "Basal-like breast cancer: molecular profiles, clinical features and survival outcomes," *BMC medical genomics*, vol. 10, no. 1, p. 19, 2017.

[6] E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *Journal of the American statistical association*, vol. 53, no. 282, pp. 457–481, 1958.

[7] T. Lucas, T. C. Silva, R. Vimieiro, and T. B. Ludermir, "A new evolutionary algorithm for mining top-k discriminative patterns in high dimensional data," *Applied Soft Computing*, vol. 59, pp. 487–499, 2017.

[8] T. Lucas, R. Vimieiro, and T. Ludermir, "SSDP+: A diverse and more informative subgroup discovery approach for high dimensional data," in *2018 IEEE Congress on Evolutionary Computation (CEC)*, pp. 1–8, IEEE, 2018.

[9] A. H. B. d. F. Lucena, "Investigação do uso de computação evolucionária para descoberta de subgrupos em conjuntos de dados numéricos de alta dimensionalidade," Master's thesis, Universidade Federal de Pernambuco, 2019.

[10] J. B. Mattos, E. G. Silva, P. S. de Mattos Neto, and R. Vimieiro, "Exceptional survival model mining," in *Brazilian Conference on Intelligent Systems*, pp. 307–321, Springer, 2020.

[11] F. Lemmerich and M. Becker, "pysubgroup: Easy-to-use subgroup discovery in python," in *Joint European conference on machine learning and knowledge discovery in databases*, pp. 658–662, Springer, 2018.

[12] Ł. Wróbel, A. Gudyś, and M. Sikora, "Learning rule sets from survival data," *BMC bioinformatics*, vol. 18, no. 1, p. 285, 2017.