



BIROn - Birkbeck Institutional Research Online

Wan, Cen and Freitas, A. (2017) An empirical evaluation of hierarchical feature selection methods for classification in Bioinformatics datasets with gene ontology-based features. *Artificial Intelligence Review* , ISSN 0269-2821.

Downloaded from: <http://eprints.bbk.ac.uk/30710/>

Usage Guidelines:

Please refer to usage guidelines at <http://eprints.bbk.ac.uk/policies.html> or alternatively contact lib-eprints@bbk.ac.uk.

An Empirical Evaluation of Hierarchical Feature Selection Methods for Classification in Bioinformatics Datasets with Gene Ontology-based Features

Cen Wan · Alex A. Freitas

Received: date / Accepted: date

Abstract Hierarchical feature selection is a new research area in machine learning/data mining, which consists of performing feature selection by exploiting dependency relationships among hierarchically structured features. This paper evaluates two hierarchical feature selection methods, used together with Naïve Bayes (NB), Tree Augmented Naïve Bayes (TAN), and Bayesian Network Augmented Naïve Bayes (BAN) classifiers. These two hierarchical feature selection methods are compared with each other and with a well-known “flat” feature selection method, i.e., Correlation-based Feature Selection (CFS). The adopted bioinformatics datasets consist of aging-related genes used as instances and Gene Ontology terms used as hierarchical features. The experimental results reveal that the HIP (Select Hierarchical Information Preserving Features) method performs best when working with three different types of Bayesian classifiers in terms of predictive accuracy and robustness on coping with data where the instances’ classes have a substantially imbalanced distribution. This paper also reports a list of the Gene Ontology terms that were most often selected by the HIP method.

Keywords Hierarchical Feature Selection · Classification · Machine Learning · Data Mining · Bayesian Classifiers · Biology of Aging

Cen Wan

Department of Computer Science, University College London, London, United Kingdom
School of Computing, University of Kent, Canterbury, United Kingdom
Fax: +44 (0)20 7387 1397
E-mail: C.Wan@ucl.ac.uk; andywan0125@gmail.com

Alex A. Freitas

School of Computing, University of Kent, Canterbury, United Kingdom
Tel.: +44 (0)1227 827220
Fax: +44 (0)1227 762811
E-mail: A.A.Freitas@kent.ac.uk

1 Introduction

In the context of the classification task of machine learning (or data mining), feature selection methods aim at improving the predictive performance of classifiers by removing redundant or irrelevant features (Liu and Motoda 1998). Feature selection is a challenging problem because the number of candidate feature subsets grows exponentially with the number of features. More precisely, the number of candidate feature subsets is $2^m - 1$, where m is the number of features. Feature selection methods can be divided into two categories (Guyon and Elisseeff 2003): embedded and pre-processing methods. Embedded methods select features during the construction of the classification model. Pre-processing methods are categorized into two groups: filter and wrapper. Filter methods select features by measuring the relevance of features regardless of the classifier, whereas wrapper methods measure the relevance of features based on the performance of a classifier. In general, filter methods are faster and more scalable than wrapper methods, so we focus on filter methods in this work.

This paper addresses a specific type of feature selection problem where the features are organized into a hierarchical structure, with more generic features representing ancestors of more specific features in the feature hierarchy. In this work the feature hierarchy is the Directed Acyclic Graph (DAG) of the Gene Ontology (GO), which, broadly speaking, contains terms specifying the hierarchical functions of genes. More precisely, in our datasets, each instance is an aging-related gene (i.e., a gene which is believed to affect the process of aging in model organisms), each feature represents a GO term (broadly speaking, a gene function) that may be present or absent for each instance (gene), and the class variable specifies whether the gene is associated with increasing or decreasing the longevity of a model organism.

Note that, although we focus on the use of a feature DAG, the methods evaluated here are also applicable to feature trees, and in general to any hierarchical feature structure where there is an “is-a” or “generalization-specialization” relationship among features, so that the presence of a feature in an instance implies the presence of all ancestors of that feature in the instance.

It is worth mentioning that the Gene Ontology is a very popular bioinformatics resource to specify gene functions, and analyzing data about aging-related genes is important because age is the greatest risk factor for a large number of age-related diseases (Tyner et al 2002; de Magalhães 2013). In addition, there are a limited number of papers reporting GO terms as a type of features used for classification. In particular, in the context of aging-related gene classification, when using GO terms and other types of features, Freitas et al. (2011) classified DNA repair genes into two categories, i.e., aging-related or non-aging related; and Fang et al. (2013) classified aging-related genes into DNA repair or non-DNA repair genes. However, such methods treated GO terms as “flat” features, ignoring their hierarchical generalization-specialization relationships.

There has been very little research so far on hierarchical feature selection methods – i.e., on feature selection methods that exploit the generalization-specialization relationships in the feature hierarchy to decide which features should be selected – for the classification task. Such hierarchical feature selection methods have been proposed in (Ristoski and Paulheim 2014; Lu et al 2013; Wang et al 2003; Jeong and Myaeng 2013). Most of these methods worked with tree-structured feature hierarchies (where a feature has at most one parent in the hierarchy) and text mining applications where instances represent documents/news and features represent words/concepts. An exception is (Lu et al 2013), where instances represent patients and features represent a tree-structured drug ontology. By contrast, in this work we address the more complex DAG-structured feature hierarchies of the GO, where a feature node can have multiple parents. Hierarchical feature selection methods have also been proposed for the task of selecting “enriched” Gene Ontology terms (terms that occur significantly more often than expected by chance) (Alexa et al 2006), which is quite different from the classification task addressed in this paper.

As far as we know, our previous work reported in (Wan and Freitas 2013; Wan et al 2015) seems to be the first work that proposed hierarchical feature selection methods to cope with the DAG-structured hierarchies of GO terms in the classification task. In that work we proposed three hierarchical feature selection methods, which were used as pre-processing methods for selecting features for the Naïve Bayes classification algorithm. In this paper we further evaluate the two best performing out of those three feature selection methods, to be reviewed in Section 3, as well as comparing with another flat feature selection method. More precisely, this current paper extends our previous work in several directions, as follows.

First, we compare two hierarchical feature selection methods proposed in (Wan and Freitas 2013; Wan et al 2015) against a well-known “flat” (non-hierarchical) feature selection method, i.e., the Correlation-based Feature Selection algorithm (Hall 1998), used as a baseline method. Second, we further evaluate the hierarchical feature selection methods following the pre-processing approach with three Bayesian network classifiers, i.e., Naïve Bayes, TAN (Tree Augmented Naïve Bayes) and BAN (Bayesian Network Augmented Naïve Bayes) classifiers, whilst in (Wan and Freitas 2013; Wan et al 2015) we used only one type of Bayesian classifier, viz., Naïve Bayes, and in (Wan and Freitas 2015) we used only a BAN classifier. Third, we evaluate the above three feature selection methods (two hierarchical methods and one flat method) on 28 datasets of aging-related genes: 4 model organisms times 7 different sets of hierarchical features. The hierarchical features used in this work involve combinations of three types of Gene Ontology terms describing gene properties (biological process, molecular function and cellular component terms), whilst the hierarchical features used in (Wan and Freitas 2013; Wan et al 2015) involve only biological process terms. In summary, to the best of our knowledge this paper is the first work to report the results of such an extensive evaluation of hierarchical feature selection methods for the classification task.

This paper is organized as follows. Section 2 briefly reviews the background about Naïve Bayes, TAN and BAN, lazy learning, Gene Ontology and hierarchical redundancy. Section 3 describes two hierarchical feature selection methods, viz. HIP and MR. Section 4 presents the experimental methodology and computational results, which are discussed in detail in Section 5. Section 6 reports the GO terms most frequently selected by the best feature selection method (HIP – Select Hierarchical Information Preserving Features). Finally, Section 7 presents a conclusion and future research directions.

2 Background

2.1 Naïve Bayes (NB) Classifier

Naïve Bayes (NB) is a well-known Bayesian classifier which is very computationally efficient and has in general good predictive performance. NB is based on the assumption that features are independent from each other, given the class variable. An example network topology is shown in Figure 1(a), where the edges indicate that each feature depends only on the class (their only parent node). To classify a testing instance, NB computes the probability of each class label c given all the feature values (x_1, x_2, \dots, x_m) of the instance using Equation (1) and assigns the instance to the class label with the greatest probability.

$$\mathbf{P}(c|x_1, x_2, \dots, x_m) \propto \mathbf{P}(c) \prod_{i=1}^m \mathbf{P}(x_i|c) \quad (1)$$

In Equation (1), m is the number of features, and the probability of a class label c given all feature values of an instance is estimated by calculating the product of the prior probability of c times the probability of each feature value x_i given c , using the above mentioned independence assumption.

2.2 Tree Augmented Naïve Bayes (TAN) Classifier

TAN is a type of semi-Naïve Bayes classifier that relaxes Naïve Bayes' feature independence assumption, by allowing each feature to depend on at most one other feature (in addition to depending on the class, which is a parent node of all features). An example network topology is shown in Figure 1(b), where all nodes except X_4 have one non-class variable parent node. This increases the ability to represent feature dependencies (which may lead to improved predictive accuracy) and still leads to reasonably efficient algorithms. TAN algorithms are not as efficient (fast) as NB, but TAN algorithms are in general much more efficient and scalable than other Bayesian classification algorithms that allow a feature to depend on several features. Among the several types of TAN algorithms, e.g., in (Friedman et al 1997; Keogh and Pazzani 1999; Jiang et al 2005; Zhang and Ling 2001), in this work we focus on one of the most

computationally efficient ones, which is based on the principle of maximizing the conditional mutual information (CMI) for each pair of features given the class attribute (Friedman et al 1997). Then, the Maximum Weight Spanning Tree (MST) will be built, where the weight of an edge is given by its CMI. Finally, one vertex of the MST is randomly selected as the root, and the edge directions are propagated accordingly.

2.3 Bayesian Network Augmented Naïve Bayes (BAN) Classifier

Comparing with Naïve Bayes and TAN classifiers, a Bayesian Network Augmented Naïve Bayes (BAN) classifier is a type of more sophisticated semi-Naïve Bayes classifier that allows each feature to have more than one parents. An example network topology is shown in Figure 1(c), where node X_4 has two parent nodes, i.e., X_1 and X_2 . The conventional algorithm to construct a BAN is analogous to the one for learning the TAN classifier (Friedman et al 1997). In this work, instead of learning the feature dependencies by conventional methods, we adopt the approach proposed in (Wan and Freitas 2013, 2015), where the network edges representing feature dependencies are simply the pre-defined edges in the feature hierarchy. More precisely, in (Wan and Freitas 2013, 2015), a classifier called GO-hierarchy-aware BAN (GO-BAN) directly adopted the edges of the Gene Ontology (GO)’s DAG (Directed Acyclic Graph) (The Gene Ontology Consortium 2000) – see Section 2.5 – as the topology of the BAN network. This has the advantages of saving computational time and exploiting the background knowledge associated with the Gene Ontology, which incorporates the expertise of a large number of biologists.

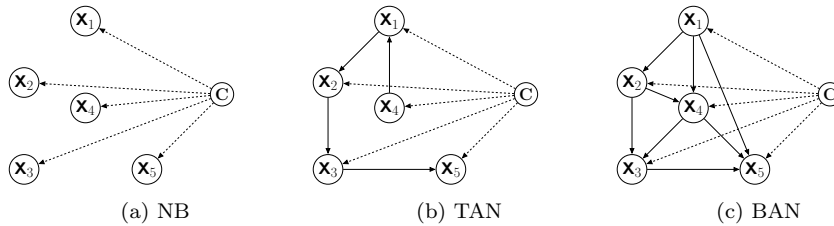


Fig. 1: Topology of Different Bayesian Network Classifiers

2.4 Lazy Learning

A “lazy” learning method performs the learning process in the testing phase, building a specific classification model for each new testing instance to be classified (Aha 1997; Pereira et al 2011). This is in contrast to the usual “eager” learning approach, where a classification model is learnt from the training

instances before any testing instance is observed. In the context of feature selection, lazy learning selects a specific set of features for each individual testing instance, whilst eager learning selects a single set of features for all testing instances. The hierarchical feature selection methods evaluated in this work are lazy methods, because they exploit hierarchical information which is specific to each instance, in order to select the best set of features for each instance – as described later. Hence, we also use lazy versions of NB, TAN and GO-BAN in our experiments.

2.5 The Gene Ontology and Hierarchical Feature Redundancy

The Gene Ontology (GO) uses unified and structured vocabularies to describe gene functions (The Gene Ontology Consortium 2000). There are three types of GO terms: biological process, molecular function and cellular component. Most GO terms are hierarchically structured by an “is-a” relationship, where each GO term is a specialization of its ancestor (more generic) terms. Therefore, there are three DAGs representing the three types of GO terms. For example, as shown in Figure 2(a), GO:0008150 (biological process) is the root of the DAG for biological process terms, and it is also the parent of GO:0051234 (establishment of localization), which is in turn the parent of GO:0006810 (transport).



(a) A small part of the Gene Ontology’s topology (b) Example of Hierarchical Redundancy

Fig. 2: Example of Hierarchically Structured Features

Consider a hierarchy of features, where each feature represents a GO term which is a node in a GO DAG. Each feature takes a binary value, “1” or “0”, indicating whether or not an instance (a gene) is annotated with the corresponding GO term. The “is-a” hierarchy of the GO is associated with two hierarchical constraints. First, if a feature takes the value “1” for a given instance, this implies its ancestors in the DAG also take the value “1” for that instance. For example, in Figure 2(a), if the term GO:0051234 has value “1” for a given gene, then the value of term GO:0008150 should be “1” as well. Conversely, if the feature takes the value “0” for a given instance, this implies

that its descendants in the DAG also take the value “0” for that instance. For example, if the term GO:0006810 has value “0”, then term GO:0044765 should also have value “0”.

Hierarchical feature redundancy is defined in this work as the case where there are two or more nodes which have the same value (“1” or “0”) in an individual instance and are located in the same path from a root to a leaf node in the DAG. For instance, in Figure 2(b), where the number “1” or “0” beside a node is the value of that feature in a given instance, nodes N and M are redundant, since both have value “1” and are located in the same path, i.e., M–N–T–K or M–N–K. Analogously, nodes T and K are redundant, since both have value “0” and are in the same path M–N–T–K. Nodes C and K are also redundant, since both have value “0” and are in the path M–C–K. Removing this type of hierarchical feature redundancy is the core task performed by the hierarchical feature selection methods used in this work.

3 Hierarchical Feature Selection Methods – Select Hierarchical Information-Preserving (HIP) Features and Select Most Relevant (MR) Features

In our previous works (Wan and Freitas 2013; Wan et al 2015; Wan and Freitas 2015), we proposed three lazy learning-based hierarchical feature selection methods to cope with the hierarchical feature redundancy issue discussed in Section 2.5. These methods are called Select Hierarchical Information-Preserving (HIP) features, Select Most Relevant (MR) features, and the hybrid HIP–MR method. In general, both HIP and MR select a set of features without hierarchical redundancy, whereas HIP–MR usually generates a set of features where the redundancy issue is only alleviated, but not eliminated (Wan et al 2015). Hence, both HIP and MR select much fewer features, and they obtained substantially greater predictive accuracy than the hybrid HIP–MR in the experiments reported in (Wan et al 2015). Hence, we use only HIP and MR in this work.

HIP and MR perform “lazy” feature selection, i.e., they select a specific set of features for each testing instance, based on the feature values observed in that instance. The HIP method selects only features whose values are not implied by the value of any other feature in the current testing instance, due to the hierarchical constraints (see Section 2.5). For instance, in Figure 3(a), the value of node (feature) C is not implied by any other feature’s value, but its value “1” implies that the values of its ancestors I, F, M, L, Q and O are also “1”; the value of node A is also not implied by any other feature’s value, but its value “0” implies that the values of nodes D, H, N, P and R are also “0”. HIP will select nodes K, B, C, A and G for the example DAG of Figure 3(a), since this feature subset preserves all the hierarchical information – i.e., for any given instance, the values of the features in this subset imply the values of all the other features.

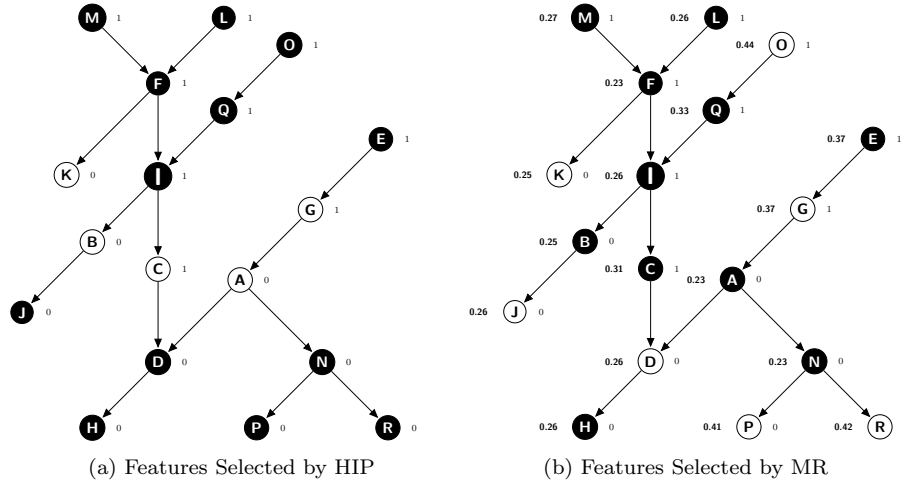


Fig. 3: Example of Hierarchical Feature Selection by HIP and MR

The MR method selects the feature with maximal relevance value in the set of features whose values equal to “1” or “0” in each path of the feature DAG. If there exist more than one features having the maximal relevance value, only the deepest (most specific) one (if the feature value is “1”) or the shallowest (most generic) one (if the feature value is “0”) in that path will be selected. Equation 2 is used to measure the relevance (\mathbf{R}) (predictive power) of a binary feature X , which can take value x_1 or x_2 . In this equation, n denotes the number of classes and c_i denotes the i -th class label. Equation 2 measures

$$\mathbf{R}(X) = \sum_{i=1}^n [\mathbf{P}(c_i|x_1) - \mathbf{P}(c_i|x_2)]^2 \quad (2)$$

the relevance of a feature as a function of the difference in the conditional probabilities of each class given different values (“1” or “0”) of the feature. In the example DAG in Figure 3(b), where the numbers on the left side of nodes denote the corresponding relevance values and the numbers on the right side of nodes denote the binary feature values, the MR method selects 7 nodes, namely K, J, D, P, R, G and O. In detail, MR selects node O rather than node Q, since the former has higher relevance value and both nodes have the value “1” in the same path; and it selects node G rather than node E, since the former is deeper than the latter and both nodes have value “1” in the same path. Analogously, MR selects node J rather than node B, since the former has higher relevance value and both nodes have the value “0” in the same path; and it selects node D rather than node H, since the former is shallower and both nodes have value “0” in the same path. Note that, in this case, the features selected by MR will lead to some hierarchical information loss. For example, the value “1” of selected node O does not imply that the value of non-selected node Q is also “1”, and the Q’s value is not implied by the

value of any selected node (so the information that Q has value “1” was lost). Similarly, the value “0” of selected node J does not imply that non-selected node B has value “0”, and B’s value is not implied by the value of any selected node.

As summarized in Table 1, both HIP and MR have the merit of eliminating hierarchical feature redundancy. However, they differ in other aspects, with their own merits and drawbacks, i.e., HIP selects features retaining all hierarchical information whilst ignoring the relevance of features with the class attribute; whereas MR selects features having higher relevance to the class attribute, but the selected features might not retain the complete hierarchical information (i.e., leading to loss of hierarchical information).

Table 1: Summary of Characteristics of the HIP and MR Methods

Hierarchical FS Algorithms	HIP	MR
Merits	<i>Eliminate hierarchical redundancy; Retain all hierarchical information</i>	<i>Eliminate hierarchical redundancy; Select highly relevant features</i>
Drawbacks	<i>Ignore relevance of features;</i>	<i>Might lead to loss of hierarchical information</i>
Feature Selectivity	<i>Select fewer features than MR</i>	<i>Select more features than HIP</i>

4 Experimental Methodology and Computational Results

4.1 Dataset Creation

We constructed 28 datasets with data about the effect of genes on an organism’s longevity, by integrating data from the Human Ageing Genomic Resources (HAGR) GenAge database (Build 17) (de Magalhães et al 2009) and the Gene Ontology (GO) database (version: 2014-06-13) (The Gene Ontology Consortium 2000). HAGR provides longevity-related gene data for four model organisms, i.e., *C. elegans* (worm), *D. melanogaster* (fly), *M. musculus* (mouse) and *S. cerevisiae* (yeast). For details of the dataset creation procedure, see (Wan and Freitas 2013; Wan et al 2015). However, in (Wan and Freitas 2013; Wan et al 2015) we created datasets using only Biological Process GO terms; whilst in this current work we created datasets with all three types of GO terms, each type associated with a hierarchy (see Section 2.5) in the form of a DAG: Biological Process (BP), Molecular Function (MF) and Cellular Component (CC) GO terms. Note that the different types of GO terms are contained in DAGs whose sets of nodes do not intersect with each other. This means that the hierarchical feature selection methods conduct the feature selection process based on each individual DAG separately.

In the created datasets, the instances represent aging-related genes, the features represent hierarchical GO terms, and the class variable indicates whether the gene contributes to increasing or decreasing the longevity of an organism. For each model organism, we created 7 datasets, with all possible subsets of the three GO term types, i.e., one dataset for each GO term type (BP, MF, CC), one dataset for each pair of GO term types (BP and MF, BP and CC, MF and CC), and one dataset with all 3 GO term types (BP, MF and CC). The structure of each created dataset is shown in Figure 4, where the feature value “1” or “0” indicates whether or not (respectively) a GO term is annotated for each gene. In the class variable, the values “Pro” and “Anti” mean “pro-longevity” and “anti-longevity”. Pro-longevity genes are those whose decreased expression (due to knock-out, mutations or RNA interference) reduces lifespan and/or whose overexpression extends lifespan. Anti-longevity genes are those whose decreased expression extends lifespan and/or whose overexpression decreases lifespan (Tacutu et al 2013).

Gene\GO	GO_1	GO_2	GO_3	...	GO_m	Class
Gene_1	1	0	0	...	0	Pro
Gene_2	0	1	0	...	1	Anti
Gene_3	0	0	0	...	1	Pro
...
Gene_n	1	0	1	...	0	Pro

Fig. 4: Structure of the Created Datasets

Note that GO terms with only one associated gene would be useless for building a classification model because they are extremely specific to an individual gene, and a model that includes these GO terms would be over-fitting the data. However, GO terms associated with only a few genes might be valuable for discovering biological knowledge, since they might represent specific biological information. In our previous work (Wan et al 2015), we did experiments with different values of a threshold defining the minimum frequency of occurrence of a GO term which is required in order to include that term (as a feature) in a dataset, in order to perform effective classification. Based on that work, the threshold value of at least 3 occurrences is used here, which retains more biological information than higher thresholds while still leading to high predictive accuracy. In addition, the root GO terms – i.e., GO:0008150, GO:0003674 and GO:0005575, respectively for the DAG of biological process, molecular function and cellular component terms – are not included in the corresponding datasets, since the root GO terms have no predictive power (all genes are trivially annotated with each root GO term).

The main characteristics of the created datasets are shown in Table 2, which reports the number of features and edges in the GO DAG, the total number of instances, the number (and percentage) of instances in each class, and the degree of class imbalance. The degree of class imbalance is calculated by Equation 3, where the degree (\mathbf{D}) equals to the complement of the ratio of the number of instances belonging to the minority class ($\mathbf{N}^o(Minor)$) over the number of instances belonging to the majority class ($\mathbf{N}^o(Major)$).

$$\mathbf{D} = 1 - \frac{\mathbf{N}^o(Minor)}{\mathbf{N}^o(Major)} \quad (3)$$

4.2 Experimental Methodology and Predictive Accuracy Measure

We evaluate two hierarchical feature selection methods – namely HIP and MR – and, as a baseline, the flat feature selection method CFS (Correlation-based Feature Selection). CFS is a well-known eager learning method that selects the same feature subset to classify all testing instances – instead of performing lazy feature selection for each testing instance, like HIP and MR. CFS tries to select a feature subset where each feature has a high correlation with the class variable and the features have a low correlation with each other (to avoid selecting redundant features). Hence, CFS is an interesting baseline method because it tries to remove redundant features in a “flat” sense, without exploiting the notion of hierarchically redundant features that is at the core of HIP and MR.

We perform three sets of experiments, using NB, TAN and GO-BAN as classifiers. The well-known 10-fold cross validation procedure was adopted to evaluate the predictive performance of NB, TAN and GO-BAN with different feature selection methods. The Geometric Mean (GMean) of the Sensitivity ($Sen.$) and Specificity ($Spe.$) is used to measure predictive accuracy, since the distributions of classes in the datasets are imbalanced. As shown in Equation 4, GMean is defined as the square root of the product of $Sen.$ and $Spe.$; $Sen.$ denotes the percentage of positive (“pro-longevity”) instances that are correctly classified as positive, whereas $Spe.$ denotes the percentage of negative (“anti-longevity”) instances that are correctly classified as negative.

$$GMean = \sqrt{Sen. \times Spe.} \quad (4)$$

4.3 Experimental Results

4.3.1 Feature selection results separately for each Bayesian classifier

Tables 3, 4 and 5 report the feature selection results separately for each type of Bayesian network classifier, namely Naïve Bayes, TAN and BAN respectively. These tables have the same structure, reporting the results for 3 different

Table 2: Main Characteristics of the Created Datasets

<i>Caenorhabditis elegans</i>							
Feature (GO term) type	BP	MF	CC	BP+MF	BP+CC	MF+CC	BP+MF+CC
N° of Features	830	218	143	1048	973	361	1191
N° of Edges	1437	259	217	1696	1654	476	1913
N° of Instances	528	279	254	553	557	432	572
N° (%) of Pro-Longevity Instances	209	121	98	213	213	170	215
	39.6%	43.4%	38.6%	38.5%	38.2%	39.4%	37.6%
N° (%) of Anti-Longevity Instances	319	158	156	340	344	262	357
	60.4%	56.6%	61.4%	61.5%	61.8%	60.6%	62.4%
Degree of Class Imbalance	0.345	0.234	0.372	0.374	0.381	0.351	0.398
<i>Drosophila melanogaster</i>							
Feature (GO term) type	BP	MF	CC	BP+MF	BP+CC	MF+CC	BP+MF+CC
N° of Features	698	130	75	828	773	205	903
N° of Edges	1190	151	101	1341	1291	252	1442
N° of Instances	127	102	90	130	128	123	130
N° (%) of Pro-Longevity Instances	91	68	62	92	91	85	92
	71.7%	66.7%	68.9%	70.8%	71.1%	69.1%	70.8%
N° (%) of Anti-Longevity Instances	36	34	28	38	37	38	38
	28.3%	33.3%	31.1%	29.2%	28.9%	30.9%	29.2%
Degree of Class Imbalance	0.604	0.500	0.548	0.587	0.593	0.553	0.587
<i>Mus musculus</i>							
Feature (GO term) type	BP	MF	CC	BP+MF	BP+CC	MF+CC	BP+MF+CC
N° of Features	1039	182	117	1221	1156	299	1338
N° of Edges	1836	205	160	2041	1996	365	2201
N° of Instances	102	98	100	102	102	102	102
N° (%) of Pro-Longevity Instances	68	65	66	68	68	68	68
	66.7%	66.3%	66.0%	66.7%	66.7%	66.7%	66.7%
N° (%) of Anti-Longevity Instances	34	33	34	34	34	34	34
	33.3%	33.7%	34.0%	33.3%	33.3%	33.3%	33.3%
Degree of Class Imbalance	0.500	0.492	0.485	0.500	0.500	0.500	0.500
<i>Saccharomyces cerevisiae</i>							
Feature (GO term) type	BP	MF	CC	BP+MF	BP+CC	MF+CC	BP+MF+CC
N° of Features	679	175	107	854	786	282	961
N° of Edges	1223	209	168	1432	1391	377	1600
N° of Instances	215	157	147	222	234	226	238
N° (%) of Pro-Longevity Instances	30	26	24	30	30	29	30
	14.0%	16.6%	16.3%	13.5%	12.8%	12.8%	12.6%
N° (%) of Anti-Longevity Instances	185	131	123	192	204	197	208
	86.0%	83.4%	83.7%	86.5%	87.2%	87.2%	87.4%
Degree of Class Imbalance	0.838	0.802	0.805	0.844	0.853	0.853	0.856

Table 3: Predictive Accuracy (%) for Naïve Bayes with Hierarchical Feature Selection Methods HIP, MR and “Flat” Feature Selection Method CFS

Feature Types	NB without Feature Selection			HIP + NB			MR + NB			CFS + NB		
<i>Caenorhabditis elegans Datasets</i>												
	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>
BP	50.2±3.6	69.0±2.6	58.9	54.1±3.4	75.5±2.8	63.9	51.2±3.5	75.5±2.6	62.2	41.1±3.3	83.7±2.6	58.7
MF	57.9±4.1	46.2±5.5	51.7	45.5±4.7	51.9±5.1	48.6	38.8±2.9	63.3±3.8	49.6	58.7±6.8	46.8±5.5	52.4
CC	43.9±5.7	70.5±3.4	55.6	58.2±4.9	60.9±4.0	59.5	42.9±4.0	71.2±3.0	55.3	35.7±4.3	74.4±3.9	51.5
BP+MF	54.0±1.8	70.3±3.0	61.6	53.5±3.6	76.2±1.9	63.8	62.9±3.5	73.2±1.8	67.9	50.2±3.5	77.1±2.4	62.2
BP+CC	52.6±3.9	68.3±2.6	59.9	57.7±3.7	73.0±2.6	64.9	55.4±2.8	73.8±2.2	63.9	44.6±3.7	77.0±2.2	58.6
MF+CC	51.2±2.8	64.1±4.3	57.3	54.7±3.3	66.0±4.1	60.1	47.6±3.6	68.3±4.2	57.0	47.1±3.9	72.1±3.8	58.3
BP+MF+CC	52.1±4.4	70.0±2.3	60.4	55.3±3.6	71.7±2.7	63.0	55.8±3.6	70.6±2.4	62.8	51.6±4.4	74.8±2.1	62.1
<i>Drosophila melanogaster Datasets</i>												
	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>
BP	74.7±3.5	36.1±9.5	51.9	73.6±4.1	44.4±9.0	57.2	79.1±4.1	38.9±11.0	55.5	76.9±4.7	27.8±7.4	46.2
MF	82.4±4.6	35.3±8.6	53.9	69.1±6.1	52.9±7.3	60.5	80.9±4.2	44.1±7.6	59.7	86.8±4.0	35.3±7.2	55.4
CC	87.1±4.1	50.0±10.2	66.0	80.6±6.5	46.4±11.4	61.2	83.9±5.6	53.6±8.7	67.1	87.1±3.3	39.3±10.0	58.5
BP+MF	77.2±3.9	50.0±10.2	62.1	72.8±5.6	57.9±9.3	64.9	79.3±4.3	44.7±8.2	59.5	85.9±3.7	31.6±7.5	52.1
BP+CC	76.9±5.1	48.6±9.8	61.1	73.6±4.9	64.9±8.3	69.1	80.2±4.3	56.8±11.2	67.5	82.4±3.7	43.2±10.9	59.7
MF+CC	89.4±3.2	57.9±5.3	71.9	82.4±6.1	63.2±6.7	72.2	83.5±4.4	57.9±7.5	69.5	91.8±3.4	42.1±8.4	62.2
BP+MF+CC	81.5±5.3	55.3±8.2	67.1	76.1±4.9	68.4±5.3	72.1	77.2±4.5	63.2±7.7	69.9	90.2±3.1	47.4±8.7	65.4
<i>Mus musculus Datasets</i>												
	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>
BP	82.4±4.7	44.1±5.9	60.3	72.1±4.8	70.6±5.1	71.3	80.9±5.2	50.0±7.9	63.6	83.8±4.0	38.2±5.6	56.6
MF	69.2±7.4	48.5±11.2	57.9	78.5±4.4	45.5±12.2	59.8	83.1±4.1	39.4±10.7	57.2	80.0±5.2	36.4±10.5	54.0
CC	75.8±2.3	52.9±10.0	63.3	80.3±3.0	47.1±11.2	61.5	81.8±3.6	41.2±11.9	58.1	71.2±3.0	35.3±11.2	50.1
BP+MF	83.8±3.4	44.1±7.0	60.8	70.6±4.8	70.6±8.1	70.6	82.4±4.2	50.0±10.2	64.2	88.2±4.2	41.2±8.0	60.3
BP+CC	79.4±6.1	50.0±8.4	63.0	66.2±5.0	73.5±9.3	69.8	73.5±5.1	52.9±9.6	62.4	83.8±5.0	50.0±11.3	64.7
MF+CC	75.0±5.0	64.7±12.5	69.7	79.4±4.2	58.8±11.8	68.3	83.8±5.0	55.9±13.3	68.4	77.9±4.8	47.1±10.9	60.6
BP+MF+CC	82.4±4.2	47.1±9.3	62.3	73.5±5.1	73.5±9.8	73.5	85.3±4.3	50.0±6.9	65.3	83.8±3.3	52.9±6.8	66.6
<i>Saccharomyces cerevisiae Datasets</i>												
	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>
BP	40.0±8.3	84.9±3.5	58.3	63.3±6.0	78.4±3.1	70.4	33.3±8.6	85.9±2.9	53.5	20.0±5.4	91.4±2.6	42.8
MF	11.5±6.1	81.7±4.8	30.7	5.0±5.0	83.2±3.4	20.4	0.0±0.0	93.9±2.4	0.0	5.0±5.0	92.4±1.8	21.5
CC	25.0±7.1	86.2±3.0	46.4	29.2±10.2	82.9±4.2	49.2	20.8±6.9	91.9±2.7	43.7	20.8±7.5	94.3±1.7	44.3
BP+MF	33.3±11.1	85.4±1.7	53.3	76.7±7.1	74.0±3.3	75.3	23.3±5.1	89.1±2.5	45.6	33.3±9.9	90.6±1.5	54.9
BP+CC	53.3±8.9	85.8±3.0	67.6	70.0±7.8	79.4±3.2	74.6	40.0±8.3	84.8±2.7	58.2	40.0±8.3	91.2±1.8	60.4
MF+CC	34.5±10.5	87.3±2.1	54.9	31.0±8.0	82.2±3.5	50.5	17.2±6.3	89.8±2.3	39.3	13.8±6.3	91.9±1.9	35.6
BP+MF+CC	36.7±9.2	85.6±2.7	56.0	70.0±10.5	75.0±2.6	72.5	30.0±9.2	86.5±2.6	50.9	36.7±10.5	92.8±1.9	58.4

Table 4: Predictive Accuracy (%) for TAN with Hierarchical Feature Selection Methods HIP, MR and “Flat” Feature Selection Method CFS

Feature Types	TAN without Feature Selection			HIP + TAN			MR + TAN			CFS + TAN		
<i>Caenorhabditis elegans Datasets</i>												
	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>
BP	34.0±3.2	79.6±2.3	52.0	52.2±2.3	67.7±3.5	59.4	55.0±2.4	73.0±1.8	63.4	45.9±3.7	79.3±2.2	60.3
MF	37.2±5.8	61.4±5.0	47.8	43.0±5.6	50.6±4.5	46.6	33.1±3.5	65.2±4.0	46.5	24.8±4.8	74.7±4.0	43.0
CC	39.8±3.0	78.2±2.2	55.8	44.9±2.7	62.2±4.7	52.8	37.8±3.4	74.4±2.7	53.0	34.7±4.3	76.9±3.2	51.7
BP+MF	35.2±1.9	80.3±2.2	53.2	54.5±3.2	72.1±2.4	62.7	61.0±4.3	71.8±2.3	66.2	46.0±3.2	80.6±2.0	60.9
BP+CC	42.7±3.1	81.7±2.7	59.1	59.2±3.9	69.2±2.9	64.0	56.3±3.0	77.3±2.2	66.0	45.1±2.8	80.8±2.0	60.4
MF+CC	40.6±3.4	74.4±3.6	55.0	45.3±2.2	67.2±3.5	55.2	45.9±3.8	70.6±3.0	56.9	47.1±3.5	73.7±3.5	58.9
BP+MF+CC	39.5±2.8	80.1±2.6	56.2	60.0±5.5	71.4±2.2	65.5	54.4±4.2	76.5±2.3	64.5	45.6±5.0	77.3±2.2	59.4
<i>Drosophila melanogaster Datasets</i>												
	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>
BP	92.3±2.9	19.4±8.4	42.3	58.2±6.5	72.2±5.4	64.8	76.9±3.6	50.0±9.6	62.0	79.1±5.1	25.0±5.9	44.5
MF	91.2±3.3	20.6±5.0	43.3	73.5±5.5	32.4±7.1	48.8	83.8±4.5	41.2±7.4	58.8	85.3±4.3	32.4±7.1	52.6
CC	90.3±3.6	32.1±11.6	53.8	79.0±3.6	50.0±11.3	62.8	75.8±6.6	42.9±8.3	57.0	87.1±3.8	42.9±10.2	61.1
BP+MF	92.4±3.3	23.7±6.9	46.8	52.2±4.0	73.7±5.8	62.0	80.4±2.8	47.4±9.5	61.7	85.9±2.9	31.6±5.3	52.1
BP+CC	86.8±4.0	18.9±7.6	40.5	59.3±5.7	67.6±7.2	63.3	82.4±3.8	40.5±8.0	57.8	79.1±5.0	48.6±10.4	62.0
MF+CC	90.6±3.3	31.6±5.0	53.5	76.5±4.9	60.5±9.3	68.0	72.9±6.4	52.6±6.9	61.9	89.4±3.8	52.6±5.8	68.6
BP+MF+CC	92.4±2.4	18.4±5.3	41.2	60.9±7.6	78.9±6.9	69.3	77.2±4.5	60.5±8.5	68.3	85.9±1.8	47.4±8.7	63.8
<i>Mus musculus Datasets</i>												
	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>
BP	89.7±3.7	41.2±4.9	60.8	42.6±5.3	73.5±7.2	56.0	73.5±7.1	50.0±10.0	60.6	82.4±3.6	47.1±6.2	62.3
MF	89.2±4.0	33.3±9.4	54.5	69.2±7.7	66.7±7.6	67.9	83.1±6.6	54.5±9.1	67.3	86.2±4.0	30.3±9.6	51.1
CC	75.8±4.4	41.2±8.3	55.9	72.7±5.1	50.0±10.1	60.3	74.2±4.3	44.1±9.8	57.2	75.8±3.2	38.2±12.6	53.8
BP+MF	86.8±3.4	35.3±5.4	55.4	42.6±4.9	79.4±9.3	58.2	79.4±4.3	55.9±8.6	66.6	88.2±4.2	41.2±8.0	60.3
BP+CC	88.2±3.6	47.1±9.7	64.5	48.5±4.4	82.4±6.8	63.2	70.6±5.9	58.8±8.9	64.4	83.8±5.0	41.2±8.7	58.8
MF+CC	88.2±4.2	41.2±10.0	60.3	63.2±3.1	64.7±12.7	63.9	82.4±3.6	55.9±11.5	67.9	77.9±3.8	52.9±10.8	64.2
BP+MF+CC	91.2±3.2	41.2±8.6	61.3	45.6±8.0	82.4±5.2	61.3	75.0±5.7	58.8±7.9	66.4	77.9±4.9	55.9±7.0	66.0
<i>Saccharomyces cerevisiae Datasets</i>												
	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>
BP	3.3±3.3	98.9±1.1	18.1	56.7±10.0	68.6±2.0	62.4	30.0±7.8	87.0±2.7	51.1	33.3±7.0	91.9±2.4	55.3
MF	0.0±0.0	97.7±1.2	0.0	26.9±6.2	78.6±2.7	46.0	0.0±0.0	87.8±2.9	0.0	5.0±5.0	94.7±1.2	21.8
CC	16.7±7.0	95.9±2.1	40.0	25.0±10.6	85.4±4.0	46.2	20.8±6.9	95.1±2.1	44.5	16.7±7.0	93.5±1.6	39.5
BP+MF	3.3±3.3	99.0±0.7	18.1	63.3±9.2	67.7±3.1	65.5	20.0±7.4	93.2±1.4	43.2	30.0±6.0	93.8±1.7	53.0
BP+CC	10.0±5.1	99.0±0.7	31.5	63.3±6.0	73.5±3.8	68.2	30.0±9.2	89.2±2.1	51.7	33.3±8.6	94.1±1.6	56.0
MF+CC	5.0±5.0	98.5±0.8	22.2	31.0±9.9	81.7±2.5	50.3	10.3±6.1	93.4±2.5	31.0	10.3±6.1	94.4±1.4	31.2
BP+MF+CC	0.0±0.0	99.0±0.6	0.0	70.0±9.2	69.7±3.0	69.8	36.7±9.2	89.4±2.1	57.3	33.3±9.9	91.8±2.1	55.3

Table 5: Predictive Accuracy (%) for GO-BAN with Hierarchical Feature Selection Methods HIP, MR and “Flat” Feature Selection Method CFS

Feature Types	GO-BAN without Feature Selection			HIP + GO-BAN			MR + GO-BAN			CFS + GO-BAN		
<i>Caenorhabditis elegans</i>												
	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>
BP	28.7 ± 2.2	86.5 ± 1.8	49.8	54.5 ± 3.2	73.4 ± 2.7	63.2	52.2 ± 3.1	74.0 ± 2.2	62.2	45.0 ± 2.6	80.9 ± 2.5	60.3
MF	34.7 ± 4.5	66.5 ± 4.5	48.0	43.8 ± 4.5	52.5 ± 5.2	48.0	35.5 ± 3.0	63.3 ± 3.4	47.4	31.4 ± 6.6	70.9 ± 6.0	47.2
CC	33.7 ± 4.5	81.4 ± 2.2	52.4	55.1 ± 5.0	63.5 ± 4.0	59.2	40.8 ± 4.3	73.1 ± 2.6	54.6	35.7 ± 4.3	74.4 ± 3.9	51.5
BP+MF	30.0 ± 2.7	84.7 ± 1.7	50.4	55.9 ± 3.2	74.1 ± 2.5	64.4	63.8 ± 2.2	73.2 ± 2.1	68.3	52.1 ± 3.7	77.6 ± 2.2	63.6
BP+CC	29.1 ± 2.1	86.6 ± 1.7	50.2	58.7 ± 3.6	72.7 ± 2.5	65.3	54.0 ± 2.8	74.7 ± 2.3	63.5	47.4 ± 2.7	79.1 ± 1.5	61.2
MF+CC	35.3 ± 2.9	80.2 ± 3.2	53.2	55.9 ± 3.1	64.5 ± 3.6	60.0	47.1 ± 3.4	70.2 ± 3.9	57.5	46.5 ± 4.1	72.1 ± 4.0	57.9
BP+MF+CC	31.2 ± 2.9	85.2 ± 1.5	51.6	58.1 ± 3.8	73.4 ± 2.6	65.3	55.3 ± 4.0	72.0 ± 2.6	63.1	50.7 ± 4.1	75.4 ± 2.1	61.8
<i>Drosophila melanogaster</i>												
	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>
BP	100.0 ± 0.0	0.0 ± 0.0	0.0	75.8 ± 4.4	52.8 ± 8.6	63.3	80.2 ± 3.5	44.4 ± 10.2	59.7	78.0 ± 4.1	25.0 ± 7.8	44.2
MF	91.2 ± 3.3	26.5 ± 3.4	49.2	64.7 ± 7.2	50.0 ± 10.0	56.9	80.9 ± 5.2	47.1 ± 9.1	61.7	85.3 ± 4.3	32.4 ± 7.1	52.6
CC	93.5 ± 2.6	28.6 ± 11.1	51.7	79.0 ± 6.6	46.4 ± 11.4	60.5	85.5 ± 4.6	42.9 ± 10.2	60.6	88.7 ± 3.5	46.4 ± 11.4	64.2
BP+MF	97.8 ± 1.5	0.0 ± 0.0	0.0	72.8 ± 3.9	63.2 ± 9.3	67.8	80.4 ± 3.7	44.7 ± 8.2	59.9	83.7 ± 3.5	28.9 ± 6.2	49.2
BP+CC	98.9 ± 1.1	0.0 ± 0.0	0.0	73.6 ± 4.7	62.2 ± 8.4	67.7	80.2 ± 4.1	51.4 ± 10.9	64.2	82.4 ± 4.4	40.5 ± 10.2	57.8
MF+CC	95.3 ± 1.9	31.6 ± 5.3	54.9	80.0 ± 6.2	60.5 ± 7.6	69.6	83.5 ± 4.9	55.3 ± 8.2	68.0	90.6 ± 3.0	52.6 ± 4.5	69.0
BP+MF+CC	98.9 ± 1.1	2.6 ± 2.5	16.0	73.9 ± 4.7	68.4 ± 5.3	71.1	81.5 ± 3.7	63.2 ± 7.7	71.8	88.0 ± 2.6	44.7 ± 8.2	62.7
<i>Mus musculus</i>												
	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>
BP	98.5 ± 1.4	26.5 ± 5.0	51.1	75.0 ± 5.1	70.6 ± 5.1	72.8	88.2 ± 4.7	44.1 ± 7.7	62.4	85.3 ± 4.3	44.1 ± 5.9	61.3
MF	90.8 ± 3.3	27.3 ± 10.0	49.8	84.6 ± 3.0	45.5 ± 12.2	62.0	87.7 ± 3.0	39.4 ± 10.6	58.8	87.7 ± 2.9	30.3 ± 9.6	51.5
CC	86.4 ± 3.3	35.3 ± 11.2	55.2	80.3 ± 3.0	50.0 ± 10.1	63.4	78.8 ± 3.8	44.1 ± 11.1	58.9	78.8 ± 3.3	38.2 ± 12.6	54.9
BP+MF	98.5 ± 1.4	29.4 ± 6.4	53.8	69.1 ± 5.8	70.6 ± 8.1	69.8	86.8 ± 4.0	41.2 ± 9.6	59.8	89.7 ± 2.2	41.2 ± 8.0	60.8
BP+CC	98.5 ± 1.4	29.4 ± 6.4	53.8	66.2 ± 6.0	76.5 ± 8.0	71.2	77.9 ± 5.3	52.9 ± 9.6	64.2	82.4 ± 5.6	47.1 ± 11.7	62.3
MF+CC	91.2 ± 3.2	26.5 ± 8.8	49.2	79.4 ± 4.2	61.8 ± 12.5	70.0	83.8 ± 5.0	58.8 ± 13.1	70.2	79.4 ± 4.8	44.1 ± 9.6	59.2
BP+MF+CC	98.5 ± 1.4	26.5 ± 10.5	51.1	70.6 ± 6.0	76.5 ± 8.8	73.5	86.8 ± 4.0	50.0 ± 6.9	65.9	83.8 ± 3.3	52.9 ± 8.4	66.6
<i>Saccharomyces cerevisiae</i>												
	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>	<i>Sen.</i>	<i>Spe.</i>	<i>GM</i>
BP	0.0 ± 0.0	100.0 ± 0.0	0.0	63.3 ± 6.0	76.8 ± 3.1	69.7	33.3 ± 8.6	89.7 ± 2.5	54.7	20.0 ± 5.4	94.6 ± 1.9	43.5
MF	0.0 ± 0.0	99.2 ± 0.8	0.0	23.1 ± 6.7	80.2 ± 3.9	43.0	0.0 ± 0.0	90.8 ± 3.0	0.0	0.0 ± 0.0	94.7 ± 1.6	0.0
CC	12.5 ± 6.1	99.2 ± 0.8	35.2	29.2 ± 10.2	83.7 ± 4.1	49.4	20.8 ± 6.9	93.5 ± 2.7	44.1	20.8 ± 7.5	93.5 ± 1.6	44.1
BP+MF	0.0 ± 0.0	100.0 ± 0.0	0.0	73.3 ± 6.7	71.9 ± 3.0	72.6	23.3 ± 7.1	89.6 ± 2.6	45.7	26.7 ± 8.3	96.4 ± 1.1	50.7
BP+CC	0.0 ± 0.0	100.0 ± 0.0	0.0	63.3 ± 10.5	78.4 ± 2.9	70.4	40.0 ± 8.3	87.3 ± 2.5	59.1	26.7 ± 6.7	96.6 ± 1.1	50.8
MF+CC	0.0 ± 0.0	100.0 ± 0.0	0.0	41.4 ± 8.3	80.7 ± 3.0	57.8	13.8 ± 6.3	88.8 ± 2.3	35.0	13.8 ± 6.3	93.4 ± 1.5	35.9
BP+MF+CC	0.0 ± 0.0	100.0 ± 0.0	0.0	76.7 ± 7.1	73.6 ± 2.8	75.1	33.3 ± 5.0	87.0 ± 2.5	53.8	23.3 ± 8.7	94.2 ± 1.6	46.8

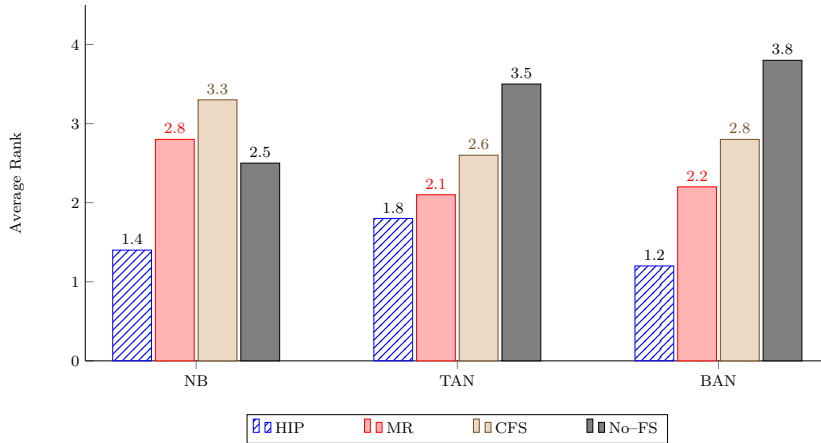


Fig. 5: Average Ranks for Feature Selection Methods with Different Classifiers

feature selection methods in a pre-processing phase, namely two hierarchical feature selection methods, HIP and MR, and the “flat” feature selection method CFS, and also when not using any feature selection method. Each table contains the results for 28 datasets – 7 combinations of feature (GO term) types for each of 4 model organisms. For each of the 28 datasets in each table, we compute a ranking of the feature selection methods, where ranks 1 and 4 represent the best and worst GMean values, respectively, in that dataset. The average rank (across the 28 datasets) is then shown in Figure 5, which summarizes the results of Tables 3–5.

Table 3 compares the predictive accuracies obtained by Naïve Bayes. As shown in the group of bars in the left side of Figure 5, the HIP method obtains the best average rank of 1.4. The second, third and worst results were obtained by Naïve Bayes without feature selection (No-FS), MR and CFS, with average ranks of 2.5, 2.8 and 3.3, respectively. It is obvious that HIP performs best when it works with Naïve Bayes, since it is ranked in the first position (indicated by a GMean value in boldface font in Table 3) in 21 out of the 28 datasets.

Table 4 compares the predictive accuracies obtained by TAN. As shown in the group of bars in the middle of Figure 5, the HIP method again obtained the best average rank (1.8), and it ranked first in 15 out of the 28 datasets in Table 4. MR obtained the second best average rank (2.1). The third and worst results were obtained by CFS and TAN without feature selection (No-FS), with average ranks of 2.6 and 3.5, respectively.

Table 5 compares the predictive accuracies obtained by GO-BAN. As shown in the group of bars in the right side of Figure 5, the HIP method again obtained the best average rank (1.2), and it ranked first in 23 out of the 28 datasets in Table 5. MR obtained the second best average rank (2.2). The third and worst results were obtained by CFS and GO-BAN without feature

selection (No-FS), with average ranks of 2.8 and 3.8, respectively.

4.3.2 Global comparison of feature selection methods combined with three different types of Bayesian classifiers

This section considers each pair of a feature selection approach combined with a type of Bayesian network classifier as a whole “classification approach”, and compares the predictive performance of the 12 classification approaches used in our experiments, rather than evaluating the results of each feature selection method separately for each type of Bayesian classifier like in the previous section. Note that we have 12 classification approaches because there are 4 feature selection approaches (3 feature selection methods and the no feature selection approach) and 3 Bayesian classifiers. Figure 6 shows the average rank (based on Gmean values) for each classification approach, across the 28 datasets. Table 6 shows the number of wins (where the highest GMean value was obtained) by each classification approach.

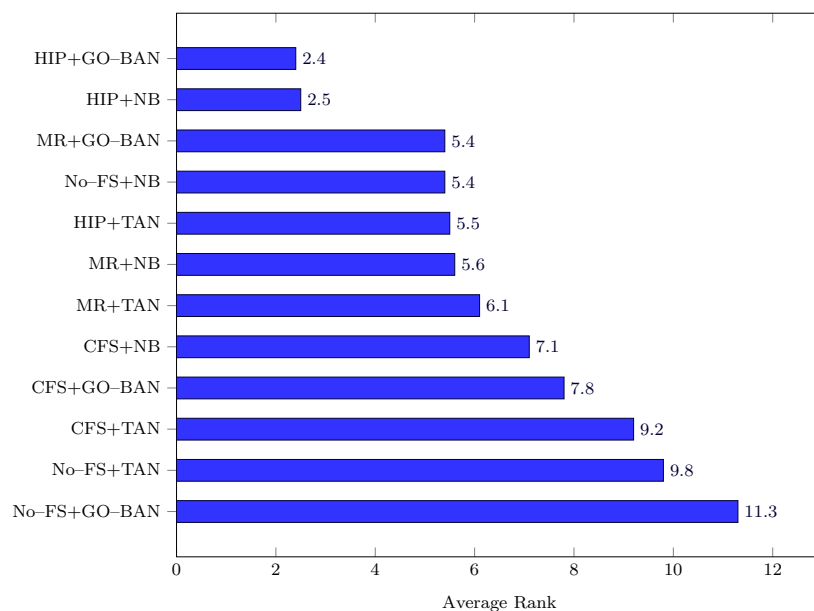


Fig. 6: Ranking of the 12 classification approaches (each combining a feature selection approach with a type of Bayesian classifier)

HIP+GO-BAN achieved the best average rank, 2.4 (Figure 6), and the second highest number of wins, 7 (Table 6). HIP+NB achieved the second best rank (2.5), and the highest number of wins, 11. Clearly both HIP+GO-BAN and HIP+NB obtained substantially better results (particularly in terms of

average rank, as shown in Figure 6 than the other classification approaches. In general, classification approaches using MR obtained better average ranks than classification approaches using CFS in Figure 6. The two worst classification approaches in that Figure involved no feature selection. In addition, as shown in Table 6, HIP and MR obtained in total 22 and 6 wins, respectively. By contrast, both CFS and the no feature selection approach did not win in any dataset, as shown in Table 6.

Table 6: Number of Wins (Best Gmean Values) Obtained by Each Combination of a Feature Selection Approach and a Bayesian Classifier

Feature Selection Methods	HIP			MR			CFS			No-FS		
	<i>NB</i>	<i>TAN</i>	<i>BAN</i>	<i>NB</i>	<i>TAN</i>	<i>BAN</i>	<i>NB</i>	<i>TAN</i>	<i>BAN</i>	<i>NB</i>	<i>TAN</i>	<i>BAN</i>
Wins	11	4	7	2	1	3	<i>0</i>	<i>0</i>	<i>0</i>	<i>0</i>	<i>0</i>	<i>0</i>
$\Sigma(\text{Wins})$	22			6			0			0		

5 Discussion

5.1 Results of Statistical Significance Tests on Predictive Accuracy

We adopted the Friedman test and Holm *post-hoc* method to conduct a significance test on the differences between the GMean values of the feature selection methods, when using NB, TAN and GO-BAN as classifiers. The Friedman test is a non-parametric statistical test based on the ranks of each classifier’s GMean value on each dataset (Japkowicz and Shah 2011; Derrac et al 2011). The Holm *post-hoc* method is used for coping with the multiple-comparison problem when using significance tests, by adjusting the significance level for pairwise method comparisons (Demsär 2006).

We firstly conducted the significance test on different feature selection methods working with each of the different Bayesian classifiers. The significance tests results are shown in Table 7, where the feature selection methods’ average ranks are shown in columns 2, 6 and 10 for those methods working with NB, TAN and GO-BAN, respectively. Recall that a lower rank value means a better predictive performance. Columns 3, 7 and 11 show the p-value result of the significance test for each feature selection method compared with the best (control) method, and columns 4, 8 and 12 show the significance level adjusted by the Holm *post-hoc* method. The difference between the ranks of two methods is deemed statistically significant when the p-value is smaller than the corresponding adjusted significance level, which is indicated by a p-value in boldface font. HIP (the best method, used as the control method when working with either NB, TAN and GO-BAN) is compared with the other feature selection methods and with not using any feature selection method (row “No-FS”).

Table 7: Statistical Test Results of the Algorithms’ GMean Values According to the Non-parametric Friedman Test with the Holm *Post-hoc* Test at the $\alpha = 0.05$ Significance Level – results for experiments in Section 4.3.1

NB				TAN				GO-BAN			
<i>FS Method</i>	<i>Ave. Rank</i>	<i>P-Value</i>	<i>Adjusted α</i>	<i>FS Method</i>	<i>Ave. Rank</i>	<i>P-Value</i>	<i>Adjusted α</i>	<i>FS Method</i>	<i>Ave. Rank</i>	<i>P-Value</i>	<i>Adjusted α</i>
HIP (ctrl.)	1.4	–	–	HIP (ctrl.)	1.8	–	–	HIP (ctrl.)	1.2	–	–
No-FS	2.5	1.4 E-03	0.0500	MR	2.1	3.8 E-01	0.0500	MR	2.2	3.7 E-03	0.0500
MR	2.8	4.9 E-05	0.0250	CFS	2.6	2.0 E-02	0.0250	CFS	2.8	3.5 E-06	0.0250
CFS	3.3	3.6 E-08	0.0167	No-FS	3.5	8.3 E-07	0.0167	No-FS	3.8	4.9 E-14	0.0167

Table 8: Statistical Test Results of the Algorithms’ GMean Values According to the Non-parametric Friedman Test with the Holm *Post-hoc* Test at the $\alpha = 0.05$ Significance Level – results for experiments in Section 4.3.2

<i>FS Method</i>	<i>Ave. Rank</i>	<i>P-Value</i>	<i>Adjusted α</i>
HIP+GO-BAN (ctrl.)	2.4	–	–
HIP+NB	2.5	9.17 E-1	5.00 E-2
MR+GO-BAN	5.4	1.85 E-3	2.50 E-2
No-FS+NB	5.4	1.85 E-3	1.67 E-2
HIP+TAN	5.5	1.30 E-3	1.25 E-2
MR+NB	5.6	0.90 E-4	1.00 E-2
MR+TAN	6.1	1.23 E-4	8.30 E-3
CFS+NB	7.1	1.08 E-6	7.10 E-3
CFS+GO-BAN	7.8	2.09 E-8	6.30 E-3
CFS+TAN	9.2	1.70 E-12	5.60 E-3
No-FS+TAN	9.8	1.60 E-14	5.00 E-3
No-FS+GO-BAN	11.3	2.56 E-20	4.60 E-3

The tests’ outcomes show that HIP significantly improves the performance of Naïve Bayes without feature selection, and significantly outperforms the MR and CFS feature selection methods when using Naïve Bayes. In addition, HIP significantly improves the performance of CFS when working with TAN, and significantly improves the predictive performance of TAN without feature selection, but it shows a non-significant difference with respect to MR using TAN. HIP also significantly outperforms MR and CFS when working with GO-BAN, and significantly improves the predictive performance of GO-BAN without feature selection.

We then conducted a significance test on the results comparing all 12 classification approaches – each of which is a combination of a feature selection approach and a type of Bayesian classifier, as explained earlier. The results are shown in Table 8, where HIP+GO-BAN is ranked first and used as the control method. The results indicate that HIP+GO-BAN significantly outperforms all other compared methods, except HIP+NB.

5.2 Number of Features Selected by HIP, MR and CFS

Figure 7 shows the average number of features selected by HIP, MR and CFS methods for each of the 7 different types of datasets, each with a different set of GO term types. Each result is the average over the 4 model organisms. HIP consistently selected fewer features than MR, while CFS selects the smallest number of features among the three methods.

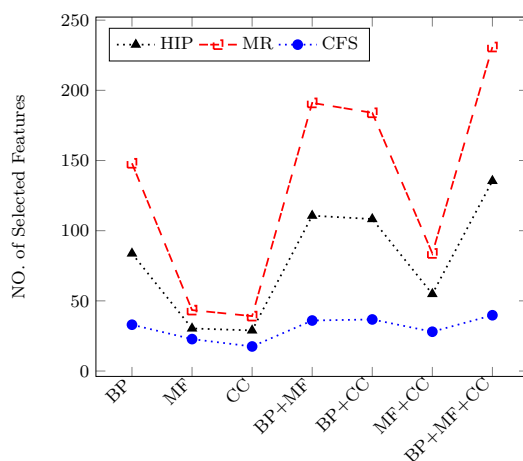


Fig. 7: Average number of features selected by HIP, MR or CFS for each of the feature (GO term) types – averaged over the 4 model organisms

5.3 Robustness of Predictive Performance Against Imbalanced Class Distributions

As shown in Figure 8, the degree of class imbalance (calculated by Equation 3) for the datasets range from 0.35, for the *Caenorhabditis elegans* datasets, to 0.84, for the *Saccharomyces cerevisiae* datasets.

We evaluated the HIP, MR and CFS methods from the perspective of robustness of predictive performance against large degrees of class imbalance, by computing the linear correlation coefficient (r) between the degree of class

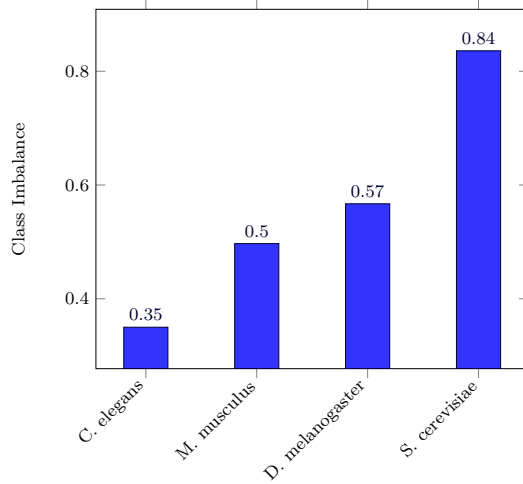


Fig. 8: Average degree of class imbalance for each of the 4 model organisms datasets – averaged over the 7 feature (GO term) types

Table 9: Correlation Coefficient between GMean and Degree of Class Imbalance for Different Feature Selection Methods with Different Classifiers

Classifier	No Feature Selection	HIP	MR	CFS
NB	-0.258	-0.035	-0.483	-0.461
TAN	-0.801	0.088	-0.515	-0.406
GO-BAN	-0.789	0.103	-0.463	-0.525

imbalance and GMean values over the 28 datasets. As shown in Table 9, HIP has the strongest robustness, since the value of its correlation coefficient is closer to “0” for Naïve Bayes, TAN and GO-BAN classifiers. The other feature selection methods, as well as the approach of no feature selection, have in general substantial negative values of the correlation coefficient, which means that their predictive accuracy tends to decrease substantially with an increase on the degree of class imbalance.

The fact that HIP is much more robust than MR and CFS to class imbalance seems to contribute substantially to HIP’s better GMean results, as explained next. First of all, note that in general HIP, MR and CFS tend to achieve higher accuracy in the prediction of majority class instances than in the prediction of minority class instances. This can be seen by noting the following two general patterns (although there are exceptions) in Tables 3, 4 and 5. First, HIP, MR and CFS exhibit in general substantially larger Specificity (Spe.) than Sensitivity (Sen.) for *C. elegans* and *S. cerevisiae* datasets, where

Table 10: Correlation Coefficient between the Degree of Class Imbalance and the Difference between Sensitivity and Specificity for Different Feature Selection Methods with Different Classifiers

<i>Classifier</i>	No Feature Selection	HIP	MR	CFS
NB	0.793	0.332	0.790	0.841
TAN	0.946	0.208	0.798	0.723
GO-BAN	0.884	0.292	0.786	0.869

Spe. measures the accuracy in the prediction of instances of the majority class (“anti-longevity” in these datasets). Second, HIP, MR and CFS exhibit in general substantially larger Sen. than Spe. for *D. melanogaster* and *M. musculus* datasets, where Sen. measures the accuracy in the prediction of instances of the majority class (“pro-longevity” in these datasets).

Next, to quantify the imbalance between Sen. and Spe. obtained by each method (HIP, MR and CFS), we computed the difference between these two terms as given by Equation 5, where *Max* and *Min* return the maximum and minimum among their two arguments, respectively. Equation 5 returns a positive value that is proportional to the difference (“imbalance”) between Sen. and Spe. Recall that $GMean = \sqrt{Sen. \times Spe.}$, which means that in order to maximize GMean one has to find a balance between maximizing both Sen. and Spe., rather than maximizing one at the expenses of minimizing the other. Then, we further calculated the linear correlation coefficient (r) between **Diff** and the degree of class imbalance given by Equation 3, as shown in Table 10. In this table, it is clear that both MR and CFS have a large positive r value, varying from 0.723 to 0.869. This means that, for these two methods, a higher degree of class imbalance tends to lead to a large **Diff** value. This tendency is much weaker for HIP, whose r values in Table 10 are much smaller, between 0.208 and 0.332. This means that HIP tends to obtain more balanced Sen. and Spe. values, leading to higher GMean values than MR and CFS, as observed in Tables 3, 4 and 5 in general.

$$\mathbf{Diff} = \mathit{Max}(\mathit{Sen}, \mathit{Spe}) - \mathit{Min}(\mathit{Sen}, \mathit{Spe}) \quad (5)$$

5.4 Summary of the Empirical Comparisons between Hierarchical Feature Selection Methods

Overall, HIP significantly outperformed both MR and CFS when working with Naïve Bayes and GO-BAN. Although HIP showed no significant difference with respect to MR when working with TAN, HIP also obtained the best average rank with that classifier. In addition, HIP showed the strongest robustness against class imbalance, working with Naïve Bayes, TAN and GO-BAN.

In our previous work in (Wan et al 2015), using only biological process GO terms as features, there was no statistically significant difference between HIP and MR when using the Naïve Bayes classifier. In this work we conclude that HIP, which eliminates hierarchical redundancy and selects the features that preserve the complete hierarchical information, performed significantly better than MR and CFS when working with Naive Bayes and GO-BAN, when coping with diverse types of GO terms – viz., biological process, molecular function and cellular component, and different combinations of these types.

6 Identifying the GO Terms (Features) Most Often Used for Classification

As the HIP method performed best with all three types of Bayesian classifiers, we computed the ranks of GO terms selected by HIP in the BP+MF+CC datasets, for each of the 4 model organisms. The top-ranked terms are shown in Table 11. The first four columns of this table have self-explanatory names. The rank in column 5 is based on two criteria. The first one is the “Frequency of Selection” in column 6, which means the number of times the GO term was selected by HIP for classifying the testing instances. The second, tie-breaking ranking criterion is the “Frequency in Edges” in column 7, which means the number of edges containing the GO term in the trees built by TAN for classifying the test instances. Recall that, for building the tree, each feature is allowed to have at most one parent feature, but each feature could be the parent for more than one child features. Hence, a feature could act as a “hub” node, if that feature is the parent for many nodes. Note that the value of “Frequency in Edges” will always be not smaller than the value of “Frequency of Selection”, since one selected feature should be included in at least one edge.

As shown in Table 11, several GO terms were very often selected across three model organisms: Synapse (GO:0045202), Extracellular Region (GO:0005576), and Antioxidant Activity (GO:0016209) are top-ranked terms in the worm, fly and mouse datasets. Other GO terms were selected across two model organisms: Reproduction (GO:0000003) and Electron Carrier Activity (GO:0009055) are top-ranked in the worm and fly datasets; Protein Binding Transcription Factor Activity (GO:0000988) in the worm and yeast datasets; Receptor Activity (GO:0004872) and Enzyme Regulator Activity (GO:0030234) in the fly and yeast datasets.

Briefly, several of these very often selected GO terms fit well with some aging-related hypotheses. For example, oxidative processes produce byproducts, i.e., ROS (reactive oxygen species), which can cause damage and crosslink DNA (Vijg and Campisi 2008); and antioxidant activity, which can mitigate the harmful effects of high-levels of ROS and is also related to the hypothesis that calorie restriction can delay aging, was found to be able to extend the longevity of model organisms like worms, mice and flies (Walker et al 2005; Wood et al 2004; Sohal and Weindruch 1996; Sohal et al 1994). As another example, in terms of the link between reproduction and aging, in *C. elegans*,

Table 11: Information about the GO Terms Most Frequently Selected by the HIP Method

Model Organism	GO Term ID	GO Term Type	GO Term Name	Rank	Freq. of Selection	Freq. in Edges	Predicted Class
<i>Worm</i>	GO:0045202	CC	synapse	1	572	2394	Anti
	GO:0000003	BP	reproduction	2	572	1929	Anti
	GO:0005576	CC	extracellular region	3	572	1095	Anti
	GO:0016209	MF	antioxidant activity	4	572	697	Pro
	GO:0040007	BP	growth	5	572	633	Pro
	GO:0022610	BP	biological adhesion	6	568	1046	Pro
	GO:0000988	MF	protein binding transcription factor activity	7	567	801	Pro
	GO:0009055	MF	electron carrier activity	8	567	779	Anti
	GO:0031974	CC	membrane-enclosed lumen	9	567	769	Anti
<i>Fly</i>	GO:0009055	MF	electron carrier activity	1	130	199	Pro
	GO:0005576	CC	extracellular region	2	130	193	Pro
	GO:0000003	BP	reproduction	3	130	184	Anti
	GO:0044456	CC	synapse part	4	130	174	Pro
	GO:0045202	CC	synapse	5	130	152	Pro
	GO:0016209	MF	antioxidant activity	6	127	354	Pro
	GO:0005198	MF	structural molecule activity	7	127	180	Pro
	GO:0030234	MF	enzyme regulator activity	8	126	144	Anti
	GO:0004872	MF	receptor activity	9	125	189	Anti
<i>Mouse</i>	GO:0044456	CC	synapse part	1	102	354	Anti
	GO:0005198	MF	structural molecule activity	2	102	344	Pro
	GO:0005576	CC	extracellular region	3	102	270	Pro
	GO:0005623	CC	cell	4	102	191	Anti
	GO:0045202	CC	synapse	5	102	124	Anti
	GO:0030054	CC	cell junction	6	99	248	Anti
	GO:0016209	MF	antioxidant activity	7	99	246	Pro
	GO:0023052	BP	signaling	8	99	207	Pro
	GO:0031012	CC	extracellular matrix	9	99	176	Pro
<i>Yeast</i>	GO:0005085	MF	guanyl-nucleotide exchange factor activity	1	238	358	Anti
	GO:0004872	MF	receptor activity	2	238	282	Anti
	GO:0022414	BP	reproductive process	3	234	511	Anti
	GO:0009295	CC	nucleoid	4	234	321	Anti
	GO:0005933	CC	cellular bud	5	231	479	Anti
	GO:0000988	MF	protein binding transcription factor activity	6	231	340	Anti
	GO:0005622	CC	intracellular	7	231	283	Anti
	GO:0032126	CC	eisosome	8	231	243	Anti
	GO:0030234	MF	enzyme regulator activity	9	230	403	Anti

mutations in the *daf-2* gene reduce insulin/insulin-like growth factor-1 (IGF-1) signaling and lead to extended lifespan and delayed reproduction (Kenyon 2010).

7 Conclusions

In summary, we evaluated the predictive performance of two hierarchical feature selection methods and compared them with the well-known “flat” feature selection method CFS (Correlation-based Feature Selection), by using Naïve Bayes, Tree Augmented Naïve Bayes and Bayesian Network Augmented Naïve Bayes classifiers over 28 aging-related gene datasets where hierarchies of Gene Ontology (GO) terms were used as predictive features. The experimental results showed that in general the HIP method performed best in terms of predictive accuracy, and it showed more robustness against a large degree of class imbalance than the other feature selection methods. We further computed the ranking of GO terms based on how often they were selected by the HIP method for classifying test instances, and identified GO terms that are among the top-ranked terms for more than one model organisms. An interesting future research direction would be to propose new hierarchical feature selection methods for coping with classification datasets where the features are non-binary – e.g., real-valued features.

Acknowledgments

We thank Dr. João Pedro de Magalhães for his valuable general advice on the biology of aging for this project. We also acknowledge the support of concurrency researchers at Kent for access to the ‘CoSMoS’ cluster, funded by EPSRC grants EP/E049419/1 and EP/E0535/1.

References

- Aha DW (1997) *Lazy Learning*. Kluwer Academic Publishers, Norwell, MA
- Alexa A, Rahnenführer J, Lengauer T (2006) Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* 22(13):1600–1607
- de Magalhães JP (2013) How ageing processes influence cancer. *Nature Reviews Cancer* 13(5):357–365
- de Magalhães JP, Budovsky A, Lehmann G, Costa J, Li Y, Fraifeld V, Church GM (2009) The human ageing genomic resources: online databases and tools for biogerontologists. *Aging Cell* 8(1):65–72
- Demsär J (2006) Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine Learning Research* 7:1–30
- Derrac J, Garcia S, Molina D, Herrera F (2011) A practical tutorial on the use of non-parametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. *Swarm and Evolutionary Computation* 1(1):3–18

- Fang Y, Wang X, Michaelis EK, Fang J (2013) Classifying aging genes into DNA repair or non-DNA repair-related categories. *Lecture Notes in Intelligent Computing Theories and Technology* pp 20–29
- Freitas AA, Vasieva O, de Magalhães JP (2011) A data mining approach for classifying DNA repair genes into ageing-related or non-ageing-related. *BMC Genomics* 12(27):1–11
- Friedman N, Geiger D, Goldszmidt M (1997) Bayesian network classifiers. *Machine Learning* 29(2-3):131–163
- Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. *The Journal of Machine Learning Research* 3:1157–1182
- Hall MA (1998) Correlation-based feature subset selection for machine learning. PhD thesis, University of Waikato, Hamilton, New Zealand
- Japkowicz N, Shah M (2011) Evaluating learning algorithms: a classification perspective. Cambridge University Press, New York, USA
- Jenatton R, Audibert JY, Bach F (2011) Structured variable selection with sparsity-inducing norms. *Journal of Machine Learning Research* 12:2777–2824
- Jeong Y, Myaeng S (2013) Feature selection using a semantic hierarchy for event recognition and type classification. In: *Proceedings of Sixth International Joint Conference on Natural Language*, Nagoya, Japan, pp 136–144
- Jiang L, Zhang H, Cai Z, Su J (2005) Learning tree augmented naive bayes for ranking. *Database Systems for Advanced Applications* pp 688–698
- Kenyon CJ (2010) The genetics of ageing. *Nature* 464(7288):504–512
- Keogh EJ, Pazzani MJ (1999) Learning augmented bayesian classifiers: A comparison of distribution-based and classification-based approaches. In: *Proc. the seventh international workshop on artificial intelligence and statistics*, Florida, USA, pp 225–230
- Liu H, Motoda H (1998) Feature selection for knowledge discovery and data mining. Kluwer Academic Publishers, Norwell, MA
- Lu S, Ye Y, Tsui R, Su H, Rexit R, Wesaratchakit S, Liu X, Hwa R (2013) Domain ontology-based feature reduction for high dimensional drug data and its application to 30-day heart failure readmission prediction. In: *Proceedings of the Ninth International Conference on Collaborative Computing: Networking, Applications and Workshar-ing (Collaboratecom)*, Austin, USA, pp 478–484
- Martins AFT, Smith NA, Aguiar PMQ, Figueiredo MAT (2011) Structured sparsity in structured prediction. In: *Proc. the 2011 conference on empirical methods in natural language processing (EMNLP 2011)*, Edinburgh, UK, pp 1500–1511
- Pereira RB, Plastino A, Zadrozny B, de C Merschmann LH, Freitas AA (2011) Lazy attribute selection: Choosing attributes at classification time. *Intelligent Data Analysis* 15(5):715–732
- Ristoski P, Paulheim H (2014) Feature selection in hierarchical feature spaces. In: *Proceedings of Seventeenth International Conference on Discovery Science*, Bled, Slovenia, pp 288–300
- Sohal RS, Weindruch R (1996) Oxidative stress, caloric restriction, and aging. *Science* 273(5271):59–63
- Sohal RS, Ku HH, Agarwal S, Forster MJ, Lal H (1994) Oxidative damage, mitochondrial oxidant generation and antioxidant defenses during aging and in response to food restriction in the mouse. *Mechanisms of ageing and development* 74(1-2):121–133
- Tacutu R, Craig T, Budovsky A, Wuttke D, Lehmann G, Taranukha D, Costa J, Fraifeld VE, de Magalhães JP (2013) Human ageing genomic resources: Integrated databases and tools for the biology and genetics of ageing. *Nucleic Acids Research* 41(D1):D1027–D1033
- The Gene Ontology Consortium (2000) Gene Ontology: tool for the unification of biology. *Nature Genetics* 25(1):25–29
- Tyner SD, Venkatchalam S, Choi J, Jones S, Ghebranious N, Igelmann H, Lu X, Soron G, Cooper B, Brayton C, Park SH, Thompson T, Karsenty G, Bradley A, Donehower LA (2002) *p53* mutant mice that display early ageing-associated phenotypes. *Nature* 415(6867):45–53
- Vijg J, Campisi J (2008) Puzzles, promises and a cure for ageing. *Nature* 454(7208):1065–1071

- Walker G, Houthoofd K, Vanfleteren JR, Gems D (2005) Dietary restriction in *C. elegans*: from rate-of-living effects to nutrient sensing pathways. *Mechanisms of ageing and development* 126(9):929–937
- Wan C, Freitas AA (2013) Prediction of the pro-longevity or anti-longevity effect of *Caenorhabditis Elegans* genes based on Bayesian classification methods. In: Proc. IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2013), Shanghai, China, pp 373–380
- Wan C, Freitas AA (2015) Two methods for constructing a gene ontology-based feature selection network for a Bayesian network classifier and applications to datasets of aging-related genes. In: Proceedings of the Sixth ACM Conference on Bioinformatics, Computational Biology and Health Informatics (ACM-BCB 2015), Atlanta, USA, pp 27–36
- Wan C, Freitas AA, de Magalhães JP (2015) Predicting the pro-longevity or anti-longevity effect of model organism genes with new hierarchical feature selection methods. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 12(2):262–275
- Wang B, McKay R, Abbass H, Barlow M (2003) A comparative study for domain ontology guided feature extraction. In: Proceedings of the Twentysixth Australasian computer science conference, Adelaide, Australia, pp 69–78
- Wood JG, Rogina B, Lavu S, Howitz K, Helfand SL, Tatar M, Sinclair D (2004) Sirtuin activators mimic caloric restriction and delay ageing in metazoans. *Nature* 430:686–689
- Ye J, Liu J (2012) Sparse methods for biomedical data. *ACM SIGKDD Explorations Newsletter* 14(1):4–15
- Zhang H, Ling CX (2001) An improved learning algorithm for augmented naive bayes. *Advances in Knowledge Discovery and Data Mining* 2035:581–586
- Zhao P, Rocha G, Yu B (2009) The composite absolute penalties family for grouped and hierarchical variable selection. *The Annual of Statistics* 37(6):3468–3497