University of Louisville

ThinkIR: The University of Louisville's Institutional Repository

5-21-2019

Galaxy classification: A machine learning analysis of GAMA catalogue data

Aleke Nolte
University of Groningen

Lingyu Wang
University of Groningen, Kapteyn Astronomical Institute

Maciej Bilicki
Leiden Observatory Research Institute

Benne W. Holwerda
University of Louisville, benne.holwerda@louisville.edu

Michael Biehl
University of Groningen

Follow this and additional works at: https://ir.library.louisville.edu/faculty

Part of the Astrophysics and Astronomy Commons

Galaxy classification:
A machine learning analysis of GAMA catalogue data

Aleke Nolte[a,*], Lingyu Wang[b,c], Maciej Bilicki[d,e], Benne Holwerda[d,f],
Michael Biehl[a]

[a]*Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence,
University of Groningen, P.O. Box 407, 9700 AK Groningen, The Netherlands*
[b]*Kapteyn Astronomical Institute, University of Groningen, Landleven 12,
9747 AD Groningen, The Netherlands*
[c]*SRON Netherlands Institute for Space Research, The Netherlands*
[d]*Leiden Observatory, Leiden University, P.O. Box 9513, 2300 RA Leiden, The Netherlands*
[e]*Center for Theoretical Physics, Polish Academy of Sciences, al. Lotników 32/46, 02-668,
Warsaw, Poland*
[f]*Department of Physics and Astronomy, 102 Natural Science Building,
University of Louisville, Louisville KY 40292, USA*

Abstract

We present a machine learning analysis of five labelled galaxy catalogues from
the Galaxy And Mass Assembly (GAMA): The *SersicCatVIKING* and *SersicCatUKIDSS* catalogues containing morphological features, the *GaussFitSimple* catalogue containing spectroscopic features, the *MagPhys* catalogue including physical parameters for galaxies, and the *Lambdar* catalogue, which contains photometric measurements. Extending work previously presented at the ESANN 2018 conference – in an analysis based on Generalized Relevance Matrix Learning Vector Quantization and Random Forests – we find that neither the data from the individual catalogues nor a combined dataset based on all 5 catalogues fully supports the visual-inspection-based galaxy classification scheme employed to categorise the galaxies. In particular, only one class, the *Little Blue Spheroids*, is consistently separable from the other classes. To aid further insight into the nature of the employed visual-based classification scheme with respect to physical and morphological features, we present the galaxy parameters that are discriminative for the achieved class distinctions.

---

*Corresponding author
Email address:* `a.f.nolte@rug.nl` (Aleke Nolte)

1. Introduction

Telescope images of galaxies reveal a multitude of appearances, ranging from smooth elliptical galaxies, through disk-like galaxies with spiral arms, to more irregular shapes. The study of morphological galaxy classification plays an important role in astronomy: the frequency and spatial distribution of galaxy types provide valuable information for the understanding of galaxy formation and evolution [1, 2].

The assignment of morphological classes to observed galaxies is a task which is commonly handled by astronomers. As manual labelling of galaxies is time consuming and expert-devised classification schemes may be subject to cognitive biases, machine learning techniques have great potential to advance astronomy by: 1) investigating automatic classification strategies, and 2) by evaluating to which extent existing classification schemes are supported by the observational data.

In this work, we extend a previous analysis [3] to make a contribution along both lines by analysing several galaxy catalogues which have been annotated using a recent classification scheme proposed by Kelvin et al. [4]. In our previous study, we assessed whether this scheme is consistent with a galaxy catalogue containing 42 astronomical parameters from the *Galaxy And Mass Assembly* (GAMA, [5]) by performing both an unsupervised and a supervised analysis with prototype-based methods. We assessed whether class structure can be recovered by a clustering of the data generated by the unsupervised Self-Organizing Map (SOM) [6], and investigated if the morphological classification can be reproduced by Generalized Relevance Matrix Learning Vector Quantization (GMLVQ) [7], a powerful supervised prototype-based method [8] chosen for its capability to not only provide classification boundaries and class-representative prototypes, but also feature relevances. Finding consis-

tently negative results for the supervised and unsupervised method, namely an intermediate classification accuracy of GMLVQ of around 73% and no clear-cut agreements between galaxy classes and SOM-clustering results, we concluded the classification scheme to be not fully supported by the considered galaxy catalogue. As discussed previously [3] the hypothesised misalignment between galaxy data and classification scheme could be explained by lack of discriminative power of the employed classifiers or clustering methods, by mis-labellings of certain galaxies (a possibility already discussed in [9]), or by the absence of essential parameters in the data set. In this work, we address two of the mentioned aspects: We employ an additional established and flexible classifier, Random Forests [10] to collect evidence that the previously found moderate classification performance is not due to shortcomings of GMLVQ. Furthermore, we address the potential incompleteness of the previously analysed dataset by performing another set of supervised analyses on several additional galaxy catalogues from the GAMA survey [11], which contain a multitude of additional photometric, spectroscopic and morphological measurements.

Despite the commonly quoted abundance of data in astronomy, well-accepted benchmark datasets are not readily available in the field of galaxy classification, and only a few works analysing GAMA catalogues with machine learning methods exist. In an analysis by Sreejith et al. [9], 10 features from GAMA catalogues are hand-selected and analysed using Support Vector Machines, Decision Trees, Random Forests and a shallow Neural Network architecture. With respect to Kelvin et. al's classification scheme a maximum classification accuracy of 76.2% is reported. Turner et al. [12] perform an unsupervised analysis of five hand-selected features from GAMA catalogues using k-means clustering. While not the main aim of Turner et al.'s analysis, a comparison of the determined clusters with class information from Kelvin et al. shows galaxies that are assigned the same class by Kelvin et al. spread over several clusters (Figures 11, 13, 15 and 17 in [12]).

In agreement with our previous results and the analyses from the above mentioned literature, we find the employed classification scheme to not be fully

3

supported even when considering the additional catalogues and an alternative classifier. Interestingly, analogous to our previous work [3], the *Little Blue Spheroids*, a galaxy class newly introduced in [4], remains most clearly pronounced, also for the set of catalogues analysed in this work. We present the parameters that are the most relevant for the achieved class distinctions.

The paper is organised in as follows: In Section 2 the analysed galaxy catalogues and their preprocessing is described. Section 3 outlines the employed classification methods, GMLVQ and Random Forests. Section 4 describes experimental setups and results. The work closes with a discussion in Section 5. This paper constitutes an extension of our contribution to the 26th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN) 2018 [3]. Parts of the text have been taken over literally without explicit notice. This concerns, among others, parts of the introduction and the description of GMLVQ in Section 3.

2. Data

In this work we analyse data from five galaxy catalogues (Table 1) containing features which have been derived from spectroscopic and photometric observations, i.e. measurements of flux intensities in different wavelength bands from the Galaxy And Mass Assembly (GAMA) survey [11] for a sample of 1295 galaxies. As the catalogues contain information for different sets of galaxies, our data set consists of the set of galaxies for which a full set of features is available after balancing the relevant classes (cf. Section 2.6).
To determine this set, each catalogue is first cross-referenced with the galaxy sample analysed in our ESANN contribution [9, 3], which contains class labels for 7941 astronomical objects. The resulting subsample is further preprocessed by selecting measurements based on the specifics of each catalogue. Subsequently, missing values are treated by first removing feature dimensions with a considerable amount of missing values (more than 500 missing values per feature

4

catalogue	shorthand	number of samples after preprocessing
GaussFitSimple	GFS	7430 galaxies with 59 emission line features
Lambdar	Lambdar	7365 galaxies with 28 flux measurements and uncertainties for different bands
MagPhys	MagPhys	7541 galaxies with 171 features
SersicCatVIKING	Viking	5476 galaxies with 66 Sérsic features
SersicCatUKIDSS	Ukidss	3008 samples with 53 Sérsic features
Complete information from all catalogues		2117 galaxies
Final sample (cf. Section 2.6)		1295 galaxies

Table 1: Overview of galaxy catalogues analysed in this work. Shown are also the number of samples for which complete information, i.e. information from each of the catalogues, is available, and the number of samples in the final dataset considered in the remainder of this work.

dimension) and then discarding samples which contain missing values in any of the remaining feature dimensions.

Details of each catalogue as well as specific processing steps are delineated in the following paragraphs.

2.1. GaussFitSimple

The GaussFitSimple catalogue (GFS) [13] contains parameters of Gaussian fits to 12 important emission lines found in galaxy spectra, namely the emission lines of oxygen ([O I] emission lines at 6300Å and 6364Å, in the following denoted as *OIB* and *OIR*, [O II] lines at 3726Å and 3729Å, denoted as *OIIB* and *OIIR*, [O III] lines at 4959Å and 5007Å, denoted as *OIIIR* and *OIIIB*), nitrogen ([N II] lines at 6548Å and 6583Å, *NIIR* and *NIIB*), sulphur ([S II] lines at 6716Å and 6731Å, *SIIR* and *SIIB*), and hydrogen (Hα and Hβ lines at 6563Å and 4861Å, respectively). Further, the catalogue contains slope and intercept of the continuum, that is, the background radiation in-between emission lines.

5

In addition to these parameters the catalogue also contains meta-information concerning model fits and corresponding errors.

From the GaussFitSimple catalogue we select amplitudes (AMP_*) and sigma (SIG_*) of the Gaussian fit for each emission line, as well as calculated fluxes (*_FLUX) and equivalent widths (*_EW). Here and in the following, the asterisk * is a placeholder for the name of the corresponding emission line. We further include information about the continuum (CONT, GRAD) and the strength of the D4000 break, resulting in 59 selected features. We discard all samples for which a failure of the fitting procedure has been indicated (FITFAIL_*), and remove samples containing missing values in any of the feature dimensions. The resulting sub-catalogue then contains 7430 galaxies with 59 emission line features.

We note that the classification performance on the full catalogue, which contains model fit information and errors / measurement uncertainties is comparable to the results achieved with the reduced catalogue containing 59 features (cf. Section 4). As the selected parameters allow for a more direct interpretation in terms of emission line strengths and therefore facilitate interpretation from the astronomical perspective, we consider the reduced catalogue in the following.

2.2. Lambdar

The Lambdar catalogue [14] contains flux measurements and uncertainties for 21 bands, as measured by the LAMBDAR software [14]. When cross-referencing with the catalogue analysed in our preceding study, 400 galaxies are missing from the Lambdar catalogue. These galaxies are removed from the considered Lambdar subset and do not contribute to the ensuing missing value calculations. Columns still containing a considerable amount of missing values after this step (> 500) are excluded from the analysis. The removed columns contain parameters that include fluxes and errors in the far and near Ultraviolet (UV) (FUV_flux, FUV_fluxerr, NUV_flux, NUV_fluxerr), and fluxes and errors in the $100\mu m$ to $500\mu m$ bands (P100_flux, P100_gcfluxerr, P160_gcflux, P160_gcfluxerr, S250_gcflux, S250_gcfluxerr, S350_gcflux, S350_gcfluxerr,

S500_gcflux, and S500_gcfluxerr). After removing these, 28 features remain in the catalogue, namely fluxes and errors for u, g, r, i and z bands observed in the Sloan Digital Sky Survey (SDSS,[15]), Z, Y, J, H and K bands from VISTA Kilo-Degree Infrared Galaxy Survey (VIKING, [16]), and W1, W2, W3 and W4 bands from the Wide-field Infrared Survey Explorer (WISE, [17]). After this step, samples that are missing measurements for any of the remaining features are removed, resulting in a final sub-catalogue of 7365 galaxies with 28 features.

2.3. MagPhys

The MagPhys catalogue [18] contains physical parameters comprising information about stellar populations as well as parameters describing the interstellar medium in the galaxies. Parameters include, among others, star formation rates, star formation time-scales, information about star formation bursts, as well as the masses of stars formed in the bursts, overall stellar ages and masses, metallicities, and information about dust in the interstellar medium and in stellar birth clouds ; all this for each included galaxy. All MagPhys parameters have been derived from information provided in the Lambdar catalogue (Section 2.2) using the MAGPHYS program [18]. Due to missing values in the Lambdar catalogue, the MagPhys catalogue does not contain information for 400 of the galaxies analysed in our ESANN contribution [3]. Apart from these, there are no missing values, so that information from 177 MagPhys features is available for 7541 galaxies. However, after selecting the final sample (cf. Section 2.6) some parameters exhibit almost no variance over the considered samples: Parameters *fb17_percentile2_5, fb18_percentile2_5, fb17_percentile16, fb17_percentile50, fb17_percentile84 and fb18_percentile16*[1] are largely constant, with maximally 15 data points displaying deviations. We therefore remove these features, which results in a dimensionality of 171 for the final MagPhys sample.

---

[1]Percentiles of the likelihood distribution of parameters describing the fraction of the effective stellar mass formed in bursts over the last 10^7 and 10^8 years

Information on the MagPhys parameter shorthand notation used in the remainder can be found in [19].

2.4. Sérsic Catalogues

Three different catalogues are available which contain parameters of single-Sérsic-component fits to the 2D surface brightness distribution of galaxies in different bands [20]. The single-Sérsic-component fits have been produced with the GALFIT program [21]. The catalogues contain a parameter, GALPLAN_*, which indicates GALFIT fitting failures for each band, where the asterisk * is a placeholder for the band. GALPLAN_*=0 indicates a severe failure when fitting the surface brightness profile of the galaxy, which could not be amended by attempting a number of correction strategies. We therefore discard all samples where GALPLAN_*=0.

An additional goodness-of-fit parameter allowing to judge the quality of profile fitting is the PSFNUM_* parameter. This parameter indicates the number of prototype stars used to model the point spread function (PSF) in the galaxy image to which the surface brightness profile was fit. As indicated in the GAMA catalogue description, modelling PSFs based on less than 10 stars may result in poor PSF models, which in turn may result in poorly fitted surface brightness distributions. Accordingly, we discard all samples where the PSFNUM_* parameters have a value lower than 10.

The catalogue further contains meta-information needed to reproduce the results of the GALFIT fitting. Here we concentrate on parameters that are descriptors of galaxies as opposed to parameters describing the fitting procedure. The galaxy descriptors, all GALFIT-derived, are: GALMAG_*, the magnitude of the Sérsic model; GALRE_*, the half-light radius measured along the semi-major axis; GALINDEX_*, the Sérsic index; GALELLIP_*, the ellipticity; GALMAGERR_*, the error on magnitude; GALREERR_*, the error on the half-light radius; GALINDEXERR_*, the error on the Sérsic index; GALELLIPERR_*, the error on ellipticity; GALMAG10RE_*, the magnitude of a model truncated at $10 \times$ the half-light radius; GALMU0_*, the central surface

8

brightness; GALMUE_*, the effective surface brightness at the half-light radius; GALMUEAVG_*, the effective surface brightness within the half-light radius; and GALR90_*, the radius containing 90% of total light, measured along the semi-major axis of the galaxy.

2.4.1. SersicCatVIKING

The SersicCatVIKING [20] catalogue contains the above measurements for the VIKING bands Z, Y, J, H, and K. Based on the GALFIT failure parameter GALPLAN_*=0, 966 samples were removed from the sub-catalogue. Additional 1074 samples were removed because of PSFNUM_* < 10. After removing samples which have missing values in any of the named feature dimensions the final sub-catalogue contains 5476 galaxies with 66 Sérsic features.

2.4.2. SersicCatUKIDSS

The SersicCatUKIDSS [20] catalogue contains the above measurements for the UKIDSS [22] bands Y, J, H, K. Based on the GALFIT failure parameter GALPLAN_*=0, 2904 samples were removed from the sub-catalogue. Additional 1841 samples were removed because of PSFNUM_* < 10. After removing samples which have missing values in any of feature dimensions the final sub-catalogue contains 3008 samples with 53 Sérsic features.

2.4.3. SersicCatSDSS

For the SersicCatSDSS catalogue [20], most samples from the cross-referenced catalogue [3, 9] are discarded based on the PSFNUM and GALPLAN selection, and only 1672 samples remain. The SersicCatSDSS catalogue is therefore excluded from the analysis.

2.5. Classification Scheme

For each galaxy analysed in our ESANN contribution [3], a class label has been determined by astronomers following a visual inspection based classification scheme described by Kelvin et al. [4]. The scheme assigns galaxies to 9 classes: *Ellipticals*, *Little Blue Spheroids*, *Early-type spirals*, *Early-type barred*

9

class index	class name	corresponding Hubble type	prevalence in data set of [3, 9]
1	Ellipticals	E0-E6	11%
2	Little blue Spheroids	-	11%
3	Early-type spirals	S0, Sa	10%
4	Early-type barred spirals	SB0, SBa	1%
5	Intermediate-type spirals	Sab, Scd	15%
6	Intermediate-type barred spirals	SBab, SBcd	2%
7	Late-type spirals & Irregulars	Sd - Irr	45%
8	Artefacts	-	0.4%
9	Stars	-	0.005%

Table 2: Overview of galaxy classes in the dataset used to cross-reference the catalogues analysed in this paper. Shown are also the corresponding Hubble types, an established galaxy type descriptor in astronomy, and the class index that is used to identify classes in the remainder of the work. Gray highlights indicate the classes that are part of the final classification problems.

spirals, *Intermediate-type spirals*, *Intermediate-type barred spirals*, *Late-type spirals & Irregulars*, *Artefacts* and *Stars* (Table 2). We will refer to the classes by their class index (1-9).

As barred spirals, artefacts and stars are highly under-represented in this sample, our subsequent analysis will focus on the substantial classes, namely classes 1, 2, 3, 5 and 7.

2.6. Sample selection

To ensure a fair comparison between the catalogues, our final dataset comprises the subsample of galaxies for which a full set of measurements is available, i.e galaxies for which measurements are provided in each of the five considered catalogues. This is the case for 2117 galaxies. Considering only the substantial classes 1, 2, 3, 5 and 7, and balancing classes so that for each class the same number of samples is selected, (259, based on class 2, the class with minimum cardinality), results in a final sample of 1295 galaxies.

3. Methods: Classifiers

3.1. GMLVQ

Generalized Relevance Matrix LVQ (GMLVQ) [7, 8] is an extension of Learning Vector Quantization (LVQ) [23]. LVQ is a supervised prototype-based method, in which prototypes are annotated with a class label. The prototypes are adapted based on the label information of the training data: if the best-matching unit (BMU), the prototype closest to the data point, is of the same class as a given data point, the prototype is moved towards the data point, while in the case of a BMU with an incorrect class label, the prototype is repelled. While LVQ assesses similarities between prototypes and data points using the Euclidean distance, GMLVQ learns a distance measure that is tailored to the data, allowing it to suppress noisy feature dimensions or to emphasise distinctive features and their pair-wise combinations. GMLVQ therefore considers a generalized distance

$$d^\Lambda(\mathbf{w}, \boldsymbol{\xi}) = (\boldsymbol{\xi} - \mathbf{w})^T \Lambda \, (\boldsymbol{\xi} - \mathbf{w}) \quad \text{with } \Lambda = \Omega^T \Omega \quad \text{and} \quad \sum_i \Lambda_{ii} = 1,$$

where Λ is an $n \times n$ positive semi-definite matrix, $\boldsymbol{\xi} \in \mathbb{R}^n$ represents a feature vector and $\mathbf{w} \in \mathbb{R}^n$ is one of M prototypes. After optimisation, the diagonal of Λ will encode the learned relevance of the feature dimensions, while the off-diagonal elements encode the relevances of pair-wise feature combinations. As empirically observed and theoretically studied [24, 25] the relevance matrix after training is typically low rank and can be used, for instance, for visualisation of the data set (see Appendix A for an example).

The parameters $\{\mathbf{w}_i\}_{i=1}^M$ and Λ are optimised based on a heuristic cost function, see [7],

$$E_{\text{GMLVQ}} = \sum_{i=1}^P \mu_i^\Lambda, \text{ with } \mu_i^\Lambda = (d_J^\Lambda(\boldsymbol{\xi}_i) - d_K^\Lambda(\boldsymbol{\xi}_i))/(d_J^\Lambda(\boldsymbol{\xi}_i) + d_K^\Lambda(\boldsymbol{\xi}_i)), \quad (1)$$

where P refers to the number of training samples, $d_J^\Lambda(\boldsymbol{\xi}) = d_J^\Lambda(\mathbf{w}_J, \boldsymbol{\xi})$ denotes the distance to the closest correctly labelled prototype $\mathbf{w}_J$, and $d_K^\Lambda(\boldsymbol{\xi}) = d_K^\Lambda(\mathbf{w}_K, \boldsymbol{\xi})$ denotes the distance to the closest incorrect prototype $\mathbf{w}_K$. If the closest prototype has an incorrect label, $d_K^\Lambda(\boldsymbol{\xi}_i)$ will be smaller than $d_J^\Lambda(\boldsymbol{\xi}_i)$, hence, the

corresponding μ_i^Λ is positive. Minimisation of E_{GMLVQ} will therefore favour the correctness of nearest prototype classification. In a stochastic gradient descent procedure based on a single example the update reads

$$\mathbf{w}_{J,K} \leftarrow \mathbf{w}_{J,K} - \eta_w \partial \mu_i / \partial \mathbf{w}_{J,K} \text{ and } \Omega \leftarrow \Omega - \eta_\Omega \partial \mu_i / \partial \Omega \,. \qquad (2)$$

Derivations and full update rules can be found in [7]. In a batch gradient descent version [26], updates of the form (2) are summed over all training samples.

3.2. Random Forests

Random Forests (RF) [10] is a well-known classification and regression method that employs an ensemble of randomised Decision Trees [27]. In randomised Decision Trees, a subset of features is chosen randomly at each node. Considering only the selected features, decision thresholds are determined based on the best attainable split between classes. To combine the classifications of each tree in the ensemble, i.e. to determine the output of the Random Forest, different methods can be employed. In the scikit-learn implementation used in our experiments [28, 29] the final classification output is obtained by averaging the probabilistic prediction of each tree.

Details on the set-up of the experiments for RF as well as for GMLVQ can be found in Section 4.1.

4. Experiments

In our experiments, we assess relevances of features and discriminability between classes by training and evaluating GMLVQ for each of the five preprocessed catalogues described in Section 2. As found in previous work [3], class 2, the Little Blue Spheroids (LBS), were particularly well-distinguishable. We perform experiments for both, the full 5-class problem, trying to distinguish between galaxy classes 1, 2, 3, 5 and 7 (cf. Table 2) and a 2-class problem in which the LBS are classified against galaxies from the other four classes. In

addition to the single catalogue experiments, we also assess feature relevances and discriminability between classes for a concatenation of all catalogues, to account for possible synergies between features from different catalogues.

To allow for interpretation in the light of other classifiers, we perform the same experiments with the widely used Random Forests (RF) classifier [10] as a baseline.

4.1. Setup

We train and evaluate GMLVQ on the galaxy catalogue data using a publicly available implementation [26]. As the GMLVQ cost function is implicitly biased towards classes with larger numbers of samples, we train and evaluate the classifier on size-balanced random subsets of the five classes. For our experiments, we specify one prototype per class and run the algorithm for 100 batch gradient steps with step size adaptation as realised in [26] with default parameter settings.the We validate the algorithm by performing a class-balanced *repeated random sub-sampling validation* (see e.g. [30] for validation methods) for a total of 10 runs. Error measures and relevance profiles shown in the following correspond to averages over the 10 repetitions. For the two-class problems we also obtain and average Receiver Operator Characteristics (ROC) and the corresponding Area under the Curve (AUC) [31].

4.1.1. Setup LBS vs others

For the two-class problem, we evaluate the classifier on a subset of the full dataset (cf. Section 2.6) containing 515 samples. For this subset, we select all 259 samples from class 2, while the *others* class is made up by 256 samples consisting of 64 samples randomly selected from class 1, 3, 5, and 7 each. The remaining settings and validation procedure remain identical to the 5-class problem.

4.1.2. Random Forests

We execute experiments employing Random Forests analogous to the GMLVQ experiments, i.e. the classifier is trained on class-balanced random subsets

13

of the data and validated using repeated random sub-sampling validation. Experiments are performed using a publicly available scikit-learn implementation [28, 29] with default settings.

4.2. Classification results based on parameters from individual catalogues

A summary of classification performances for both the 5-class and the 2-class problem can be found in Figure 1. For the 5-class problem, an overview of confusion matrices (averaged over all validation runs) for each of the catalogues is shown in Figure 1a; an overview of the average classification accuracies can be found in Figure 1c in the bottom panel. For the 2-class problem, a comparison of ROC curves and classification accuracies can be found in Figure 1b and in Figure 1c in the top right subfigure, respectively. The corresponding average relevance profiles contrasting feature relevances for the 5-class and 2-class problem are shown in the Appendix, in Figure B.1 (Lambdar catalogue), Figure B.2 (GaussFitSimple catalogue), Figure B.3 (SersicCatVIKING catalogue), Figure B.4 (SersicCatUKIDSS catalogue), and Figures B.5 and B.6 (MagPhys catalogue).

Results based on SersicCatVIKING. The confusion matrix indicating the GM-LVQ class-wise accuracy on the SersicCatVIKING catalogue exhibits similar, albeit slightly worse performance than the performances presented in our previous work [3] that was based on a different set of galaxy parameters. Based on the SersicCatVIKING, the LBS are classified with higher accuracy (87% vs. 91% in ESANN) than the other classes (47-67%, 64-74%). As in the ESANN results, classes 1 and 3 show some overlap (21% of class 1 samples are classified as class 3, and 20% of class 3 samples are erroneously classified as class 1). However, unlike in the ESANN results, the overlap between class 1 and class 2 is increased in the classification using SersicCatVIKING: 22% of class 1 samples are now classified as belonging to class 2, where this overlap was only 10% for the data analysed in our ESANN contribution [3]. This is also reflected in the 2-class problem when distinguishing the LBS from the other classes. In

14
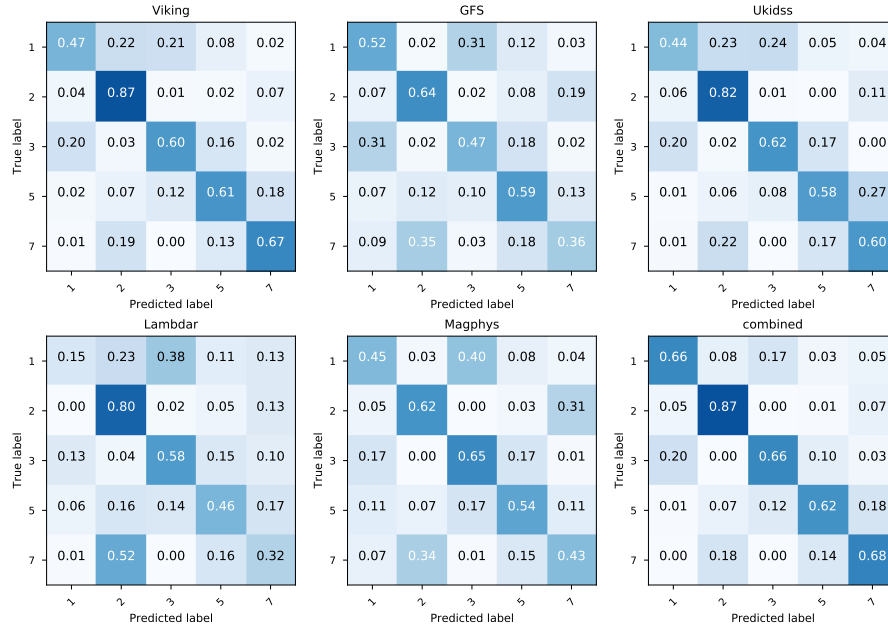
(a) 5 class problem: Confusion matrices for GMLVQ performance for each of the 5 single catalogues and the combined catalogue.
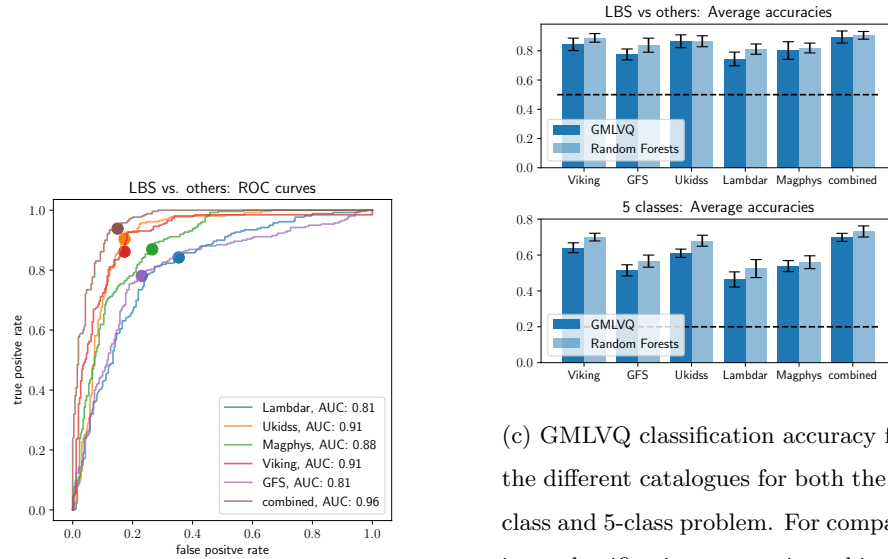
(b) 2 class problem: ROC curves for each of the 5 single catalogues and the combined catalogue.

(c) GMLVQ classification accuracy for the different catalogues for both the 2-class and 5-class problem. For comparison, classification accuracies achieved by Random forests are shown-side by side. Note that the combined condition for the Random Forests considers all available parameters, while for GMLVQ a relevance-based pre-selection is made. Dashed lines indicate chance level performance.

Figure 1: Summary of GMLVQ classification performances for both single and combined catalogues.

[3] this can be achieved with AUC(ROC)=0.96, while for the SersicCatVIKING catalogue the classification accuracy is around 84% and the AUC(ROC)=0.91. Another notable increase in overlap is the overlap between class 5 and 7, where the misclassification rate of class 5 galaxies as class 7 galaxies is increased from 8% to 18%.

Results based on GaussFitSimple Catalogue. The confusion matrix for the classification based on the GaussFitSimple Catalogue shows the highest classification accuracy of 64% for the LBS. Class 3 drops in accuracy to 47% . This is in part due to an increased overlap between the classes, 31% of class 1 samples are classified as class 3 samples and 31% of class 3 samples as belonging to class 1. In addition, there is increased overlap between class 1 and 5 (12%) and class 3 and 5 (18%), while the overlap between classes 1 and 3 with both LBS and class 7 remains low. It is notable that based on the information in the GaussFitSimple Catalogue, class 7 is only classified slighly above chance level, with most of its samples being misclassified as class 2 (35%) and class 5 (18%). Despite this, the distinction between LBS and others is still on average 78% correct, the AUC(ROC)=81%.

Results based on SersicCatUKIDSS. The results for the SersicCatUKIDSS show an overall similar performance to the results of the SersicCatVIKING catalogue: In comparison to the classification performance presented in our ESANN contribution [3], there is an increased misclassification of class 1 samples as class 2 samples, and an increased misclassification of class 5 samples as belonging to class 7. LBS classification accuracy is at 87% with an AUC(ROC)=0.91.

Results based on Lambdar Catalogue. The results for the Lambdar sample show a similar picture as the GaussFitSimple sample: Class 7 is classified with an accuracy of only slightly above chance level and is often (52%) misclassified as class 2. Unlike in the GFS results, the accuracy for class 1 is below chance level (15%). As has been the case for the other catalogues, class 1 samples are misclassified mostly as class 3 (38%). In contrast to the GaussFitSimple catalogue,

here class 1 also shows considerable overlap with class 2 (23% of class 1 samples are misclassified as class 2). In addition, a considerable amount of class 1 samples (11% and 13%) are also misclassified as classes 5 and 7. Further, class 5 and class 3 show overlap, with 15-16% misclassifications. Overall, classification accuracy based on the Lambdar catalogue is lowest (46%), while the LBS can be distinguished with 74% accuracy and an AUC(ROC)=0.81 .

Results based on MagPhys catalogue. The classification results for the MagPhys sample show a similar trend as the results based on the Lambdar sample: Classes 1 and 3 exhibit considerable overlap (40% of class 1 samples are classified as class 3, and 17% of class 3 samples are classified as class 1), class 7 accuracy is low (43%) and is frequently misclassified as class 2 (34% of the cases). In contrast to the Lambdar sample, there is almost no overlap between class 1 and class 2. Average classification accuracy for the 5 classes based on the MagPhys catalogue is at (54%), while the LBS can be distinguished with 80% accuracy and an AUC(ROC)=0.88 .

LBS vs other. The LBS can be distinguished from the other classes with an intermediate accuracy of about 74% - 87% and AUC(ROC) values of 81%-91%.

4.3. Combined catalogues

Combining all catalogues would result in a very high-dimensional classification problem, thereby rendering the resulting relevance profiles difficult to interpret. We therefore select a subset of parameters from each individual catalogue based on the feature relevances obtained in the single catalogue experiments in the following manner: For each individual catalogue, parameters are sorted according to their relevance. Subsequently, the most relevant parameters cumulatively comprising 50% of the summed total relevance are carried over to the combined catalogue. We note that we have also performed GMLVQ experiments on the full catalogue comprising all 377 features, which resulted in similar, albeit slightly worse performances than reported below.
For the Random Forests baseline experiments, we select the full catalogue of
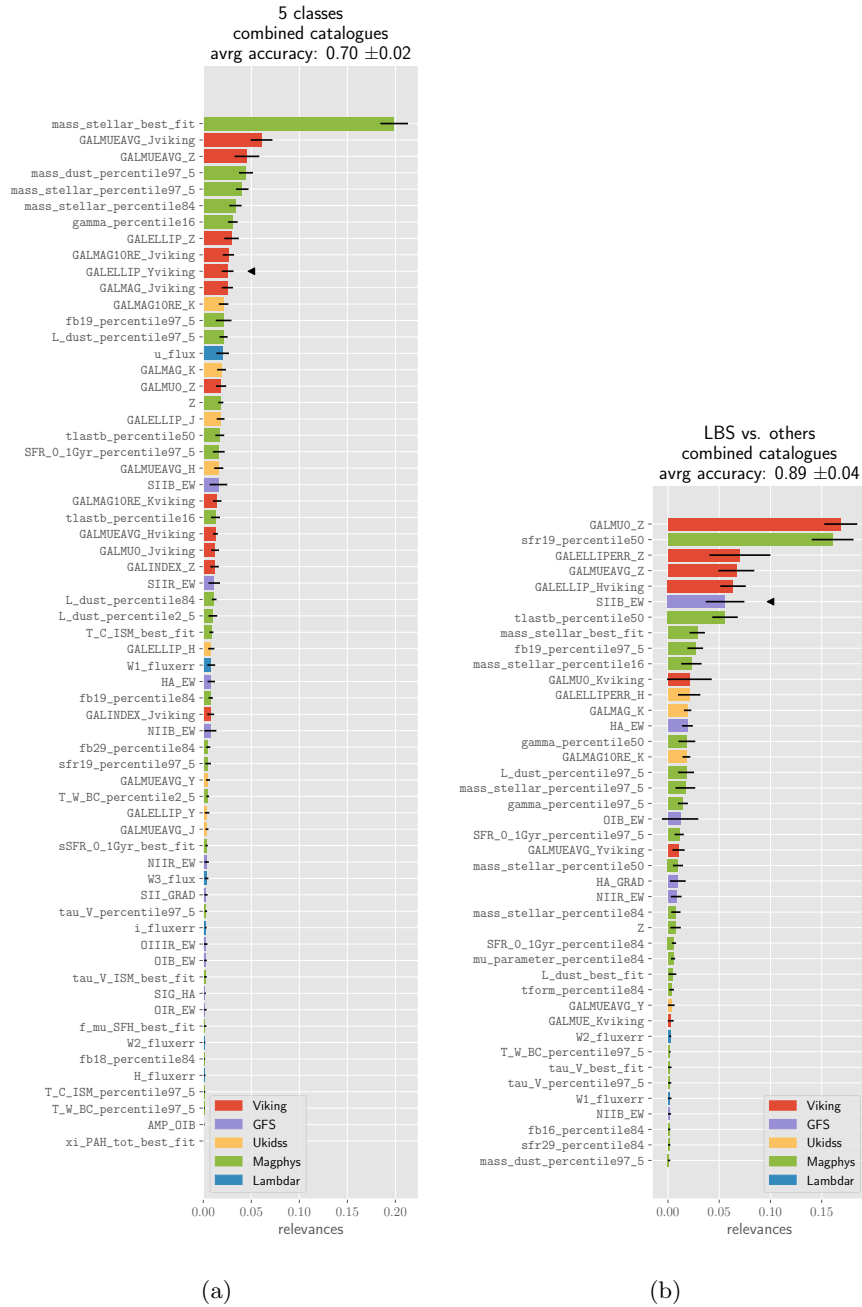
Figure 2: Sorted relevance profiles for catalogues obtained by combining the most relevant features that cumulatively make up 50% of the relevances in the single catalogue relevance profiles. Bar colours indicate the origin catalogue for each feature. Features up to the position marked by a black arrow constitute 50% of the cumulative relevance determined for the resulting combined catalogue.

377 features independent from the GMLVQ results, as to warrant identical experimental conditions. For completeness, we note that classification accuracy of Random Forests on the above described relevance-selected parameter subset is comparable to the classification accuracy on the full dataset.

Sorted relevance-profiles for the resulting combined catalogues are displayed in Figure 2a and Figure 2b, for the 5-class and 2-class problem, respectively. To simplify comparison, the confusion matrix as well as the 2-class classification performance are displayed alongside the individual catalogue performances in Figure 1.

Considering the confusion matrix for the combined catalogue, a slight overall increase in performance with respect to the individual catalogue performances can be observed. Further, it reflects the combined properties of the individual catalogues: An overlap between classes 1 and 3, some overlap between class 3 and 5, and some overlap between class 2 and 7. In comparison to the results presented in [3], classification accuracy is slightly decreased (70% vs. 73%). It should be noted however, that in [3] thrice as many samples per class were available, which could account for the difference in performance. LBS can be distinguished from the other classes with a classification accuracy of 89% and an AUC(ROC)=0.96.

Feature relevances for the combined catalogues. The parameters that make up 50% of the relevances for the 5-class and the 2-class problem (indicated by a black arrow in Figure 2a and Figure 2b), almost exclusively originate from the SersicCatVIKING and MagPhys catalogues. For the 5-class problems, these parameters are related to stellar masses and dust (mass_stellar_best_fit, mass_dust_percentile97_5, mass_stellar_percentile_97_5 and mass_stellar_percentile84), and the star formation timescale (gama_percentile16), the effective surface brightness within the half-light radius for the J- and Z-bands (GALMUEAVG_J and GAL-MUEAVG_Z), ellipticity of the galaxy (GALELLIP_Z, GALELLIP_Yviking), and magnitude of a GALFIT model of the galaxy (GALMAG10RE_Jviking).

For the 2-class problem, the most relevant parameters encompass the GALFIT central surface brightness in Z-band (GALMU0_Z), parameters related to star formation rates (sfr19_percentile50), information related to the ellipticity of the galaxies (GALELLIPERR_Z, GALELLIP_Hviking), effective surface brightness (GALMUEAVG_Z) and information about the equivalent width of the sulphur emission line.

It should be noted that relevance-matrices are not necessarily unique. They depend on which other features are available and on the parameters chosen for both data preprocessing and execution of the algorithm. This can be illustrated when considering highly correlated variables: GMLVQ might assign either two intermediate relevances to each of the variables, or deem one variable highly relevant at expense of the other correlated variable's relevance. Relevance profiles therefore should be interpreted in the sense that focusing on the most relevant parameters would allow differentiation between classes with the reported accuracy, while keeping in mind that other combinations of features may achieve this as well.

4.4. Random Forests baseline results

The classification accuracies for Random Forests for the individual and combined catalogues are displayed in Figure 1c side-by-side with the GMLVQ results. For all catalogues applying the Random Forest classifier results in comparable, though slightly better classification accuracies.

5. Discussion & Conclusion

The results presented above suggest that there may be inconsistencies in the investigated morphological classification scheme: Analogous to our previous findings [3], it has proven difficult to distinguish galaxy types using two powerful and flexible classifiers, GMLVQ and Random Forests. In all GMLVQ analyses of the individual as well as of the combined catalogues, class 1 (*Ellipticals*) and 3 (*Early-type spirals*) are particularly difficult to differentiate. Class 7 (*Late-type*

20

spirals & Irregulars) is frequently misclassified as class 5 (*Intermediate-type spirals*) and with a similar frequency as class 2 (LBS), while class 2 is consistently detected with the highest sensitivity among all classes.

The difficulty of training a successful classifier was also observed in [9], where class-wise averaged accuracies are around 75%. As mentioned in our earlier contribution [3], possible explanations for poor classification performance may be the lack of discriminative power of the employed classifiers or mis-labellings of certain galaxies [9]. A possible indication for the latter case may be that samples from class 7 (*Late-type spirals & Irregulars*) are often misclassified as class 5 (*Intermediate-type spirals*), and class 2 (LBS). This indicates that the feature representations of the galaxies in question share more properties with the named classes, and it is not unlikely that in the hand-labelling process an *Intermediate-type spiral* is occasionally misclassified as class 7 (e.g. confused with a *Late-type spiral*), or that a LBS is classified as class 7 (an *Irregular*). In the former case, employing even more flexible classifiers, e.g. GMLVQ with local relevance matrices [7], may improve classification performances. In the second case, if mis-labellings are restricted to "neighboring" classes in an assumed underlying class ordering (e.g. when considering class 5 adjacent to class 7, or class 1 (*Ellipticals*) as adjacent to class 3 (*Early-type spirals*)), ordinal classification may provide further insights [32, 33].

Despite trying to address the issue of essential parameters being not contained in the dataset analysed in [3] by considering 5 additional catalogues with a multitude of photometric, spectroscopic and morphological measurements, it is still possible that additional (and possibly not yet discovered) parameters would enable improved class distinction. Yet, our results do not rule out the possibility that the true, underlying grouping of galaxies is considerably different and less clear-cut than the investigated one. Further data-driven analyses of galaxy parameters and images with advanced clustering methods might reveal alternative groupings, like recently found for data in the VIMOS Public Extragalactic Redshift Survey [34], or even suggest novel classification schemes.

To aid further insight into the nature of the employed visual-based classification

scheme, in particular with respect to physical parameters, we have presented relevances of the catalogue features for the investigated class distinctions. Note that relevances have to be interpreted with regard to the characteristics of the data sample (e.g. correlations) and classification performance. This connotes that feature relevances are only meaningful when the class of interest is at least moderately well distinguished from the others. Further it should be noted that the presented feature relevances are not necessarily unique – alternative relevance solutions may exist. It is of particular interest to note that in the combined catalogue the most relevant features originate from the Sérsic catalogues and the MagPhys catalogue. The high relevance of Sérsic features indicate the importance of galaxy structure in different bands for the class distinction, while the presence of highly relevant features from the MagPhys catalogue highlights that classification performance is aided by these physical parameters as well. Further insight into the role of features in the context of necessary and dispensable features may be obtained by studying feature relevance bounds along the lines of [35].

Conclusions. We have presented an analysis of five galaxy catalogues using Random Forests and GMLVQ, a prototype-based classifier. Analogous to results obtained in preceding work on a lower-dimensional dataset, we conclude that even when considering a multitude of additional galaxy descriptors, the visual-based classification scheme used to label the galaxy sample remains not fully supported by the available data. Taking into account that perceptual and conceptual biases likely play non-negligible roles in the creation and application of galaxy classification schemes, further data-driven analyses might help provide novel insights regarding the true underlying grouping of galaxies.

Appendix A. Dataset visualizations and intrinsic dimensionality reduction in GMLVQ

Figures A.1 and A.2 display projections of each dataset considered in this work onto the first and second eigenvector of the relevance matrix Λ (cf. Section 3) and onto the first two principal components determined by Principal Component Analysis (PCA) [36]. The rightmost column of each figure contrasts the eigenvalue spectra of Λ and the data covariance matrix which forms the basis for PCA. While Λ is an $n \times n$ matrix, the steeply declining eigenvalue spectra for each dataset illustrate the low-dimensional subspace which GMLVQ operates in after learning [24, 25]. In particular, for the 5 class problem, Λ spans an approximately 3 dimensional subspace, while for the 2 class problem the subspace is essentially one-dimensional. The low-rank relevance matrices therefore can be thought of as performing a GMLVQ-intrinsic dimensionality reduction.

Comparing the 2-D projections onto the two leading eigenvectors of Λ and the projections onto the first two principal components, the former results in a more fanned out representation with respect the classes. This is due to the fact that by making use of the class labels, GMLVQ finds a lower-dimensional discriminative subspace as opposed to the unsupervised PCA.
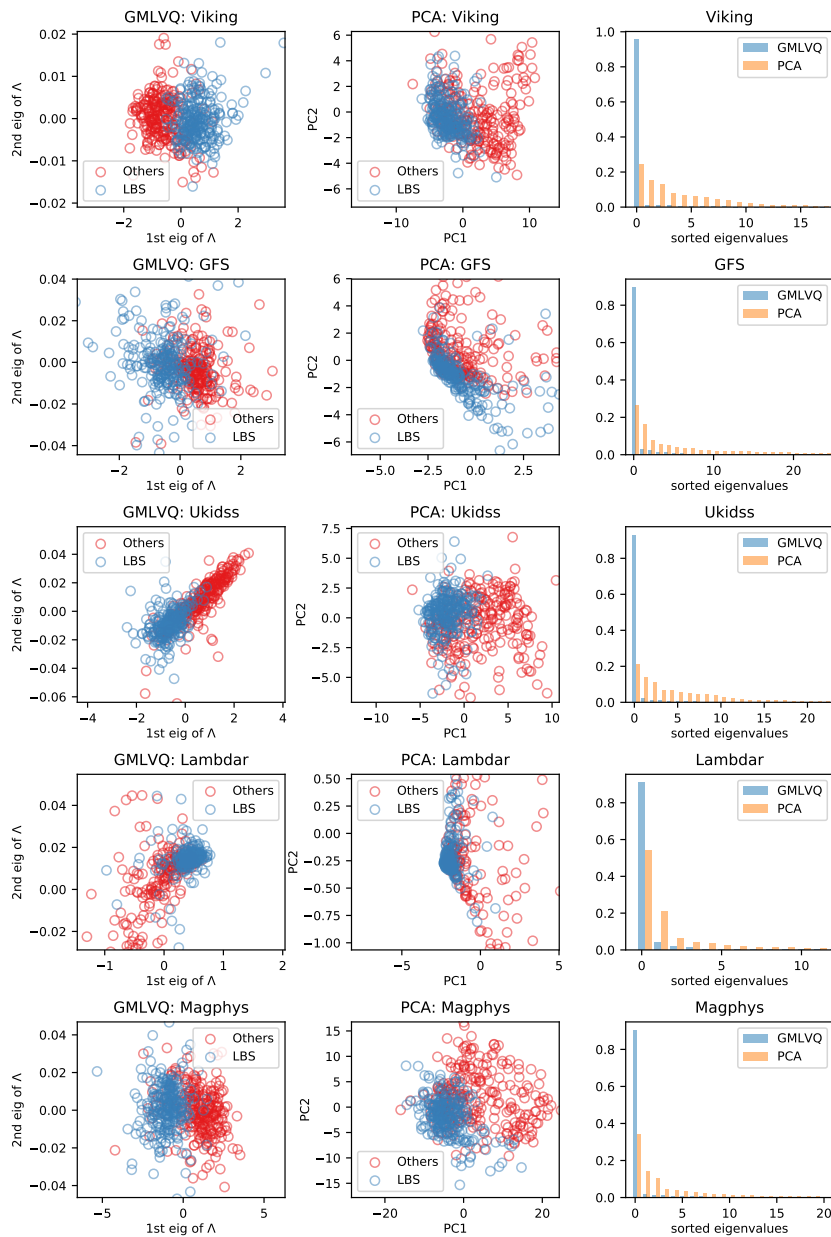
Figure A.1: 2D visualisations of the datasets used in the *LBS vs. others* classification condition. The leftmost column displays a projection of each dataset onto the first two eigenvectors of the learned relevance matrix Λ. In the middle column, projections of the datasets onto the first two principal components (PC1 and PC2) are shown. The right column juxtaposes the eigenvalue spectra of the relevance matrix and the data covariance matrix used in PCA. For increased readability, figures concentrate on the median region of the data and axes are cut off at a 3 times inter-quantile range distance from the median. Furthermore, the data projections are scaled by the square root of the corresponding eigenvalues. In the sub-figures of the eigenvalue spectra the x-axis is truncated after both eigenvalues have dropped below a value of 0.005.
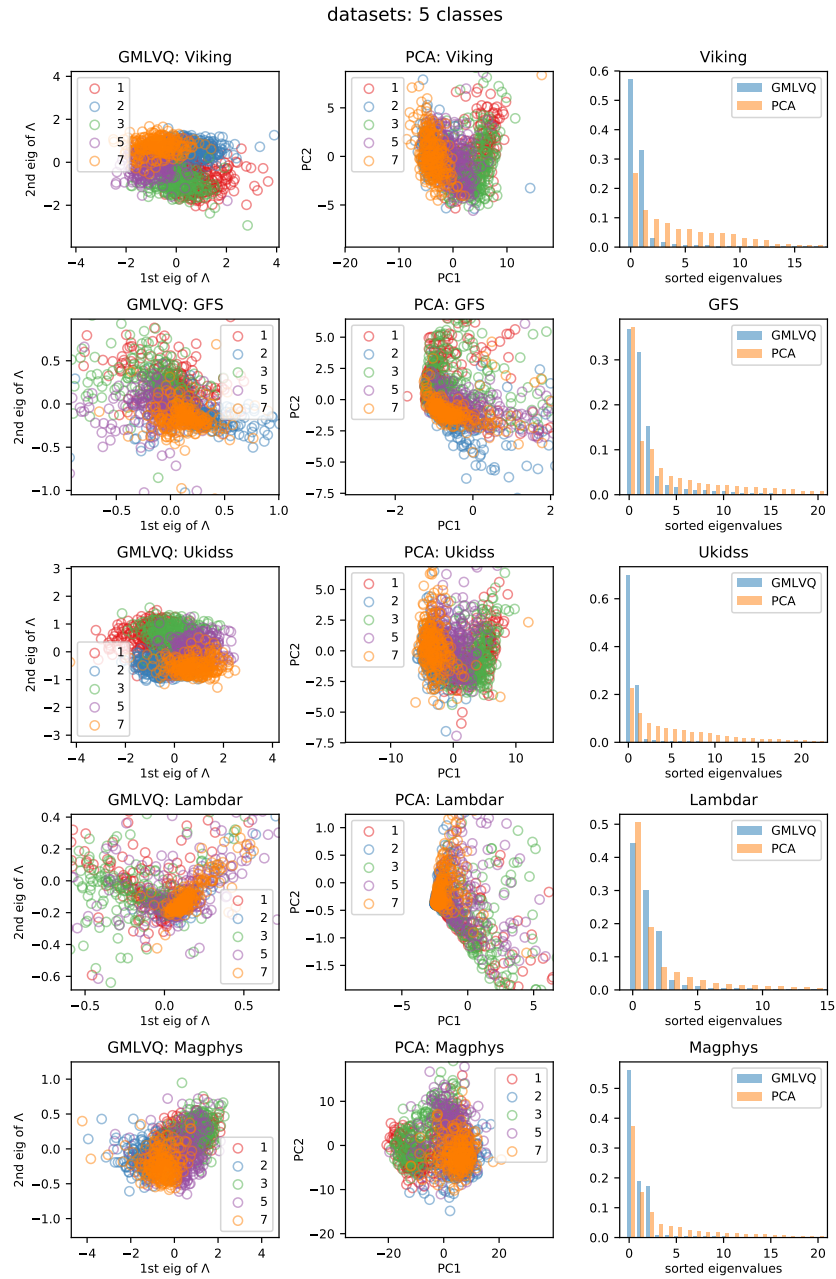
Figure A.2: 2D visualizations of the datasets used in the *5-class* classification condition. The leftmost column displays a projection of each dataset onto the first two eigenvectors of the learned relevance matrix Λ. In the middle column, projections of the datasets onto the first two principal components (PC1 and PC2) are shown. The right column juxtaposes the eigenvalue spectra of the relevance matrix and the data covariance matrix used in PCA. For increased readability, figures concentrate on the median region of the data and axes are cut off at a 3 times inter-quantile range distance from the median. Furthermore, the data projections are scaled by the square root of the corresponding eigenvalues. In the sub-figures of the eigenvalue spectra the x-axis is truncated after both eigenvalues have dropped below a value of 0.005.

Appendix B. Feature relevances for individual catalogues

In the following (Figures B.1- B.6), we present relevance profiles for the individual catalogues analysed in this work. Relevance profiles reflect the diagonal of GMLVQ's relevance matrix Λ after learning (cf. Section 3) and summarise the importance of features for a given data sample and classification task. Figures display mean and variance of the profiles over 10 independent runs (cf. Section 4.1). As noted previously, for an accurate interpretation it is important to note that, in general, relevance profiles are not unique: Especially in the presence of correlated variables, alternative profiles resulting in comparable classification performance might exist. In particular, a feature's low relevance does not entail the feature to carry no information for the desired class distinction, but may instead indicate its contribution to be at least partly redundant with other features.

For example, contrary to expectations at first glance, our experiments with the Lambdar sample result in relevance profiles that indicate uncertainties of fluxes of various bands as more relevant than the corresponding flux measurements themselves (Figure B.1). While it is not unthinkable that flux uncertainties systematically vary over a subset of galaxy classes (personal communication, Angus Wright, developer of the LAMBDAR software), in our sample W1 and W2 fluxes are correlated with both their respective errors and with fluxes from other bands. W1 and W2 fluxes as well as fluxes from other bands are thus at least partly redundant with the W1 and W2 flux uncertainties, and therefore might end up more relevant than the corresponding fluxes.
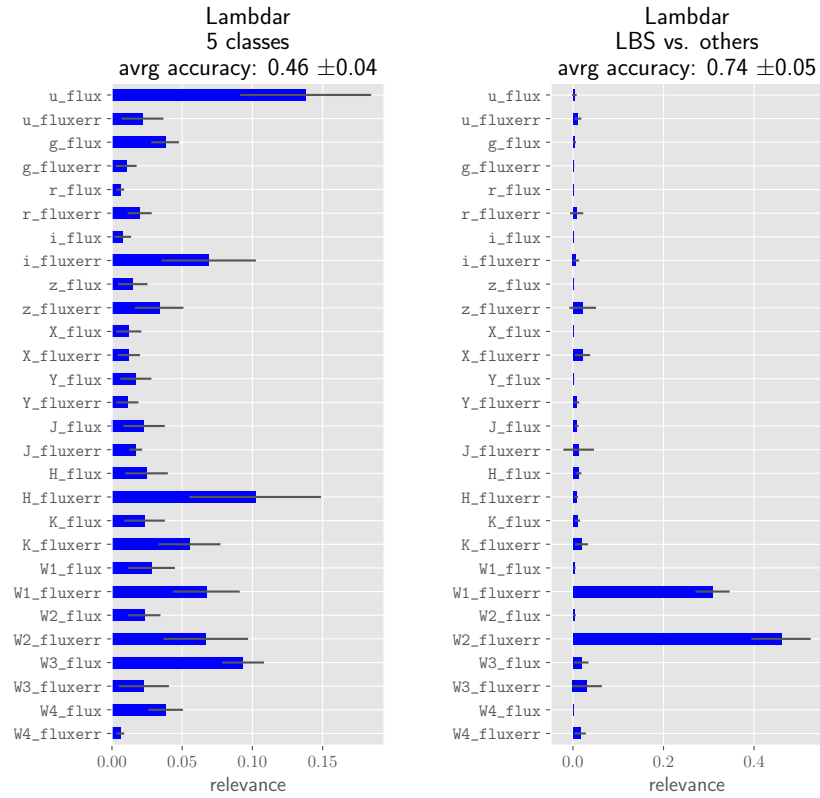
Figure B.1: Feature relevances as determined by GMLVQ for the Lambdar sample. For accurate interpretation of the relevance profiles, take note that relevance profiles are not necessarily unique, in particular in the presence of highly correlated variables. This connotes that focusing on the relevant parameters would enable to differentiate between classes with the reported accuracy, however, there may be other combinations of features which could result in similar accuracies.
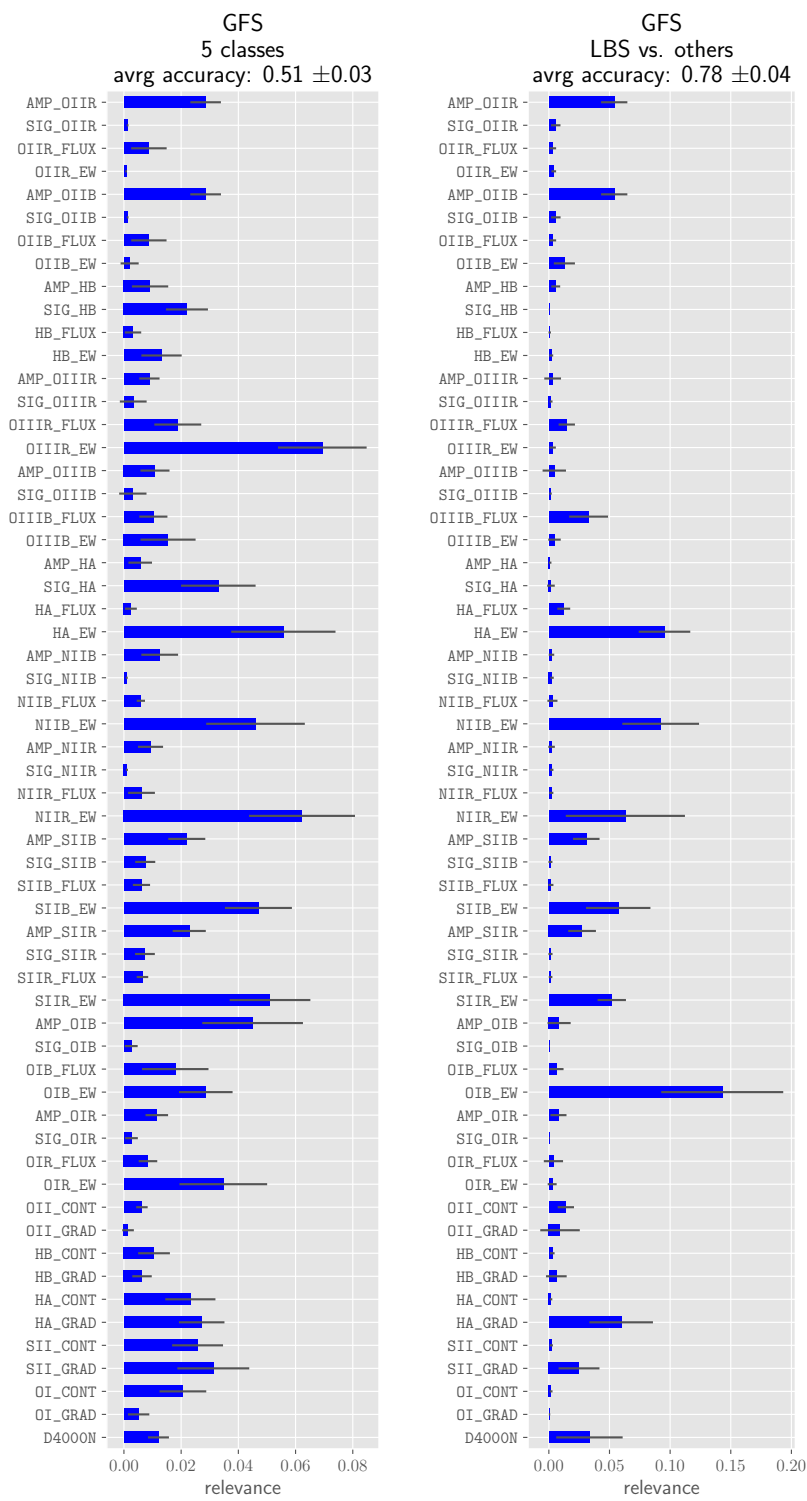
Figure B.2: Feature relevances as determined by GMLVQ for the GaussFitSimple sample. Same note applies here as to Fig. B.1.
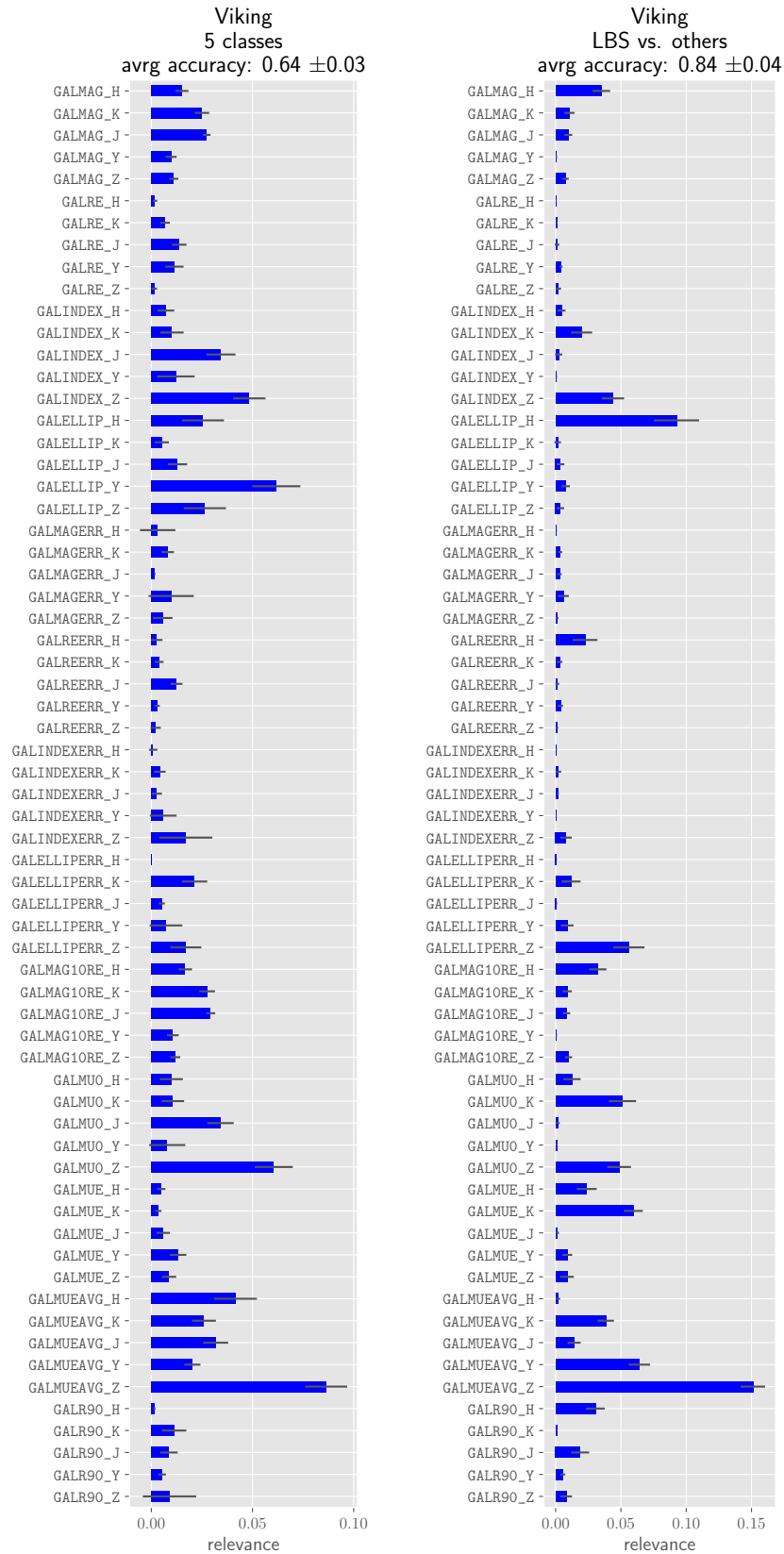
Figure B.3: Feature relevances as determined by GMLVQ for the SersicCatVIKING sample.
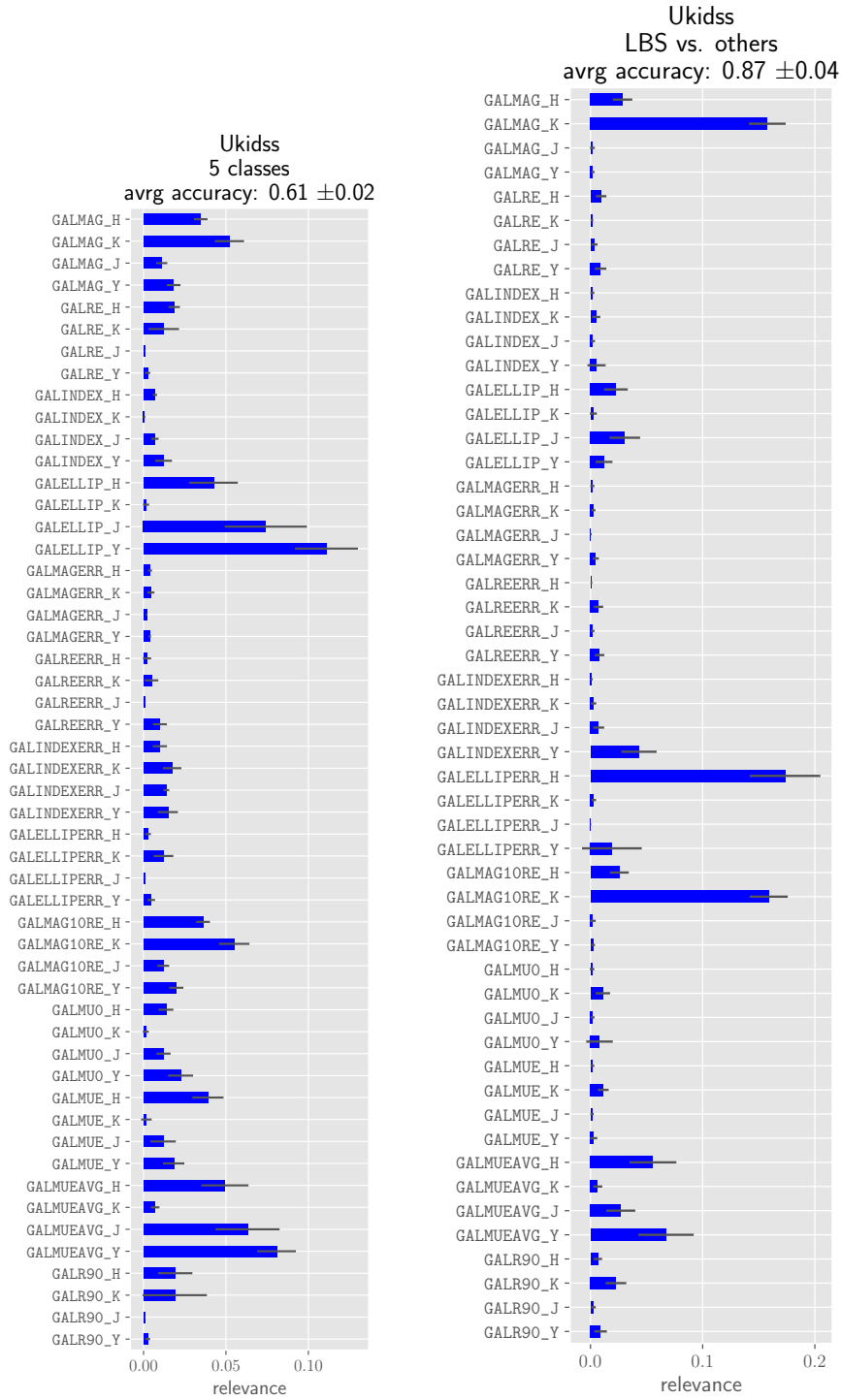
Same note applies here as to Fig. B.1.

Figure B.4: Feature relevances as determined by GMLVQ for the SersicCatUKIDSS sample.
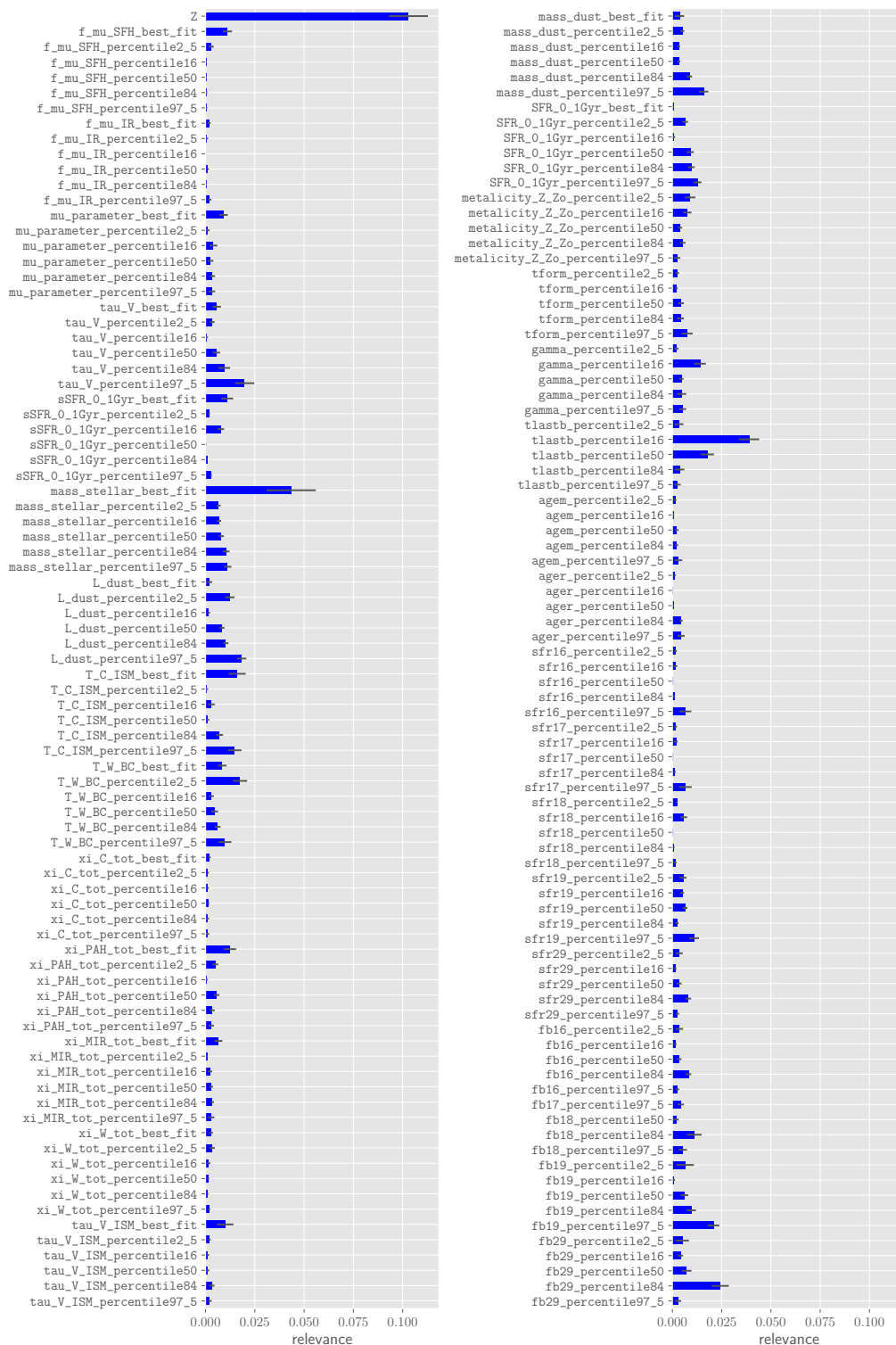Same note applies here as to Fig. B.1.

Figure B.5: 5 class problem: Feature relevances as determined by GMLVQ for the MagPhys sample. Same note applies here as to Fig. B.1.
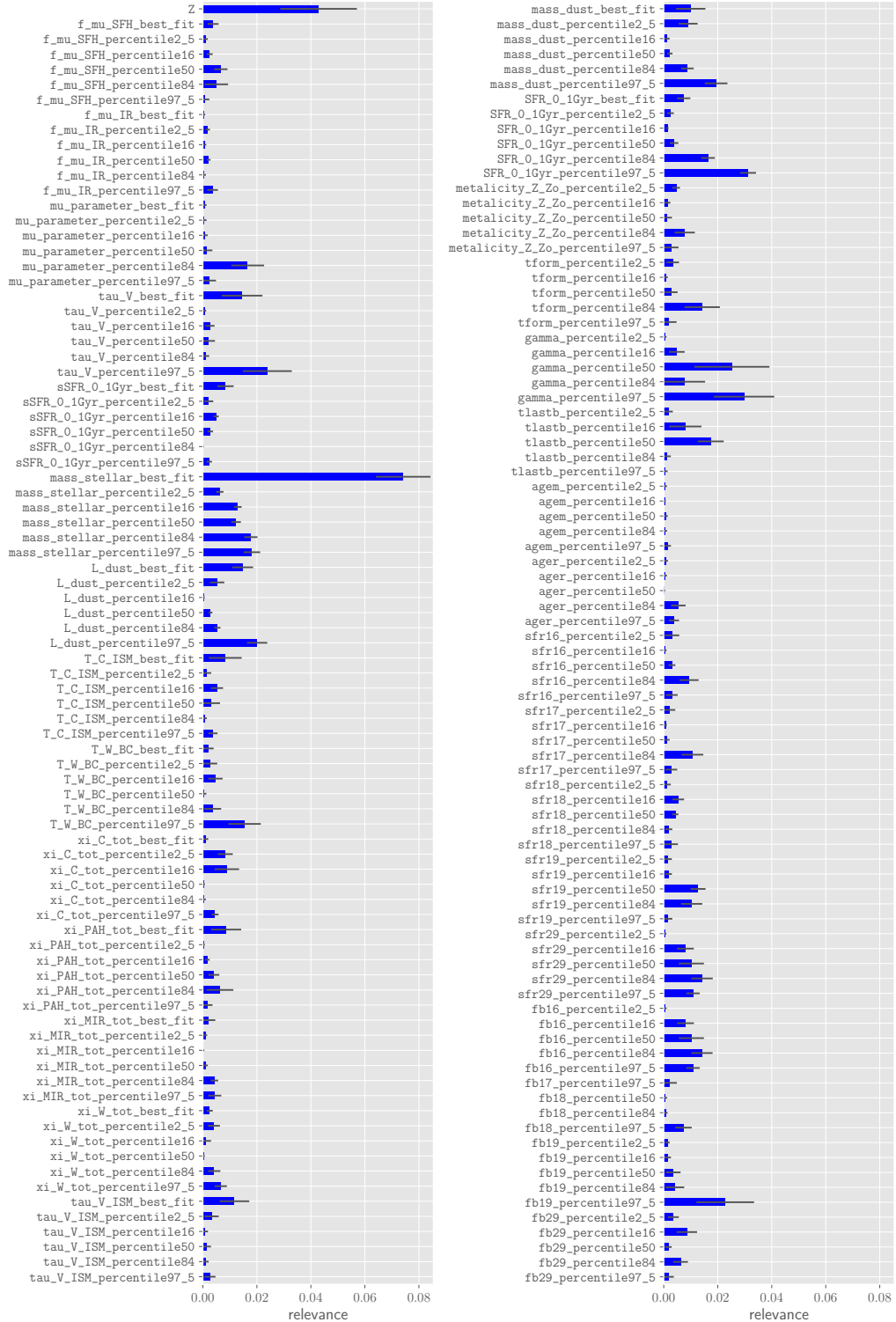
Figure B.6: LBS vs. others: Feature relevances as determined by GMLVQ for the MagPhys sample. Same note applies here as to Fig. B.1.

References

[1] R. J. Buta, Galaxy morphology, in: T. D. Oswalt, W. C. Keel (Eds.), Planets, Stars and Stellar Systems: Volume 6: Extragalactic Astronomy and Cosmology, Springer Netherlands, Dordrecht, 2013, pp. 1–89. `doi:10.1007/978-94-007-5609-0_1`.
URL `https://doi.org/10.1007/978-94-007-5609-0_1`

[2] H. Mo, F. Van den Bosch, S. White, Galaxy formation and evolution, Cambridge University Press, 2010.

[3] A. Nolte, L. Wang, M. Biehl, Prototype-based analysis of GAMA galaxy catalogue data, in: M. Verleysen (Ed.), Proceedings of the 26th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, Ciaco - i6doc.com, 2018, pp. 339–344.

[4] L. Kelvin, S. Driver, A. S. Robotham, A. W. Graham, S. Phillipps, N. K. Agius, M. Alpaslan, I. Baldry, S. Bamford, J. Bland-Hawthorn, et al., Galaxy And Mass Assembly (GAMA): ugrizYJHK Sérsic luminosity functions and the cosmic spectral energy distribution by Hubble type, Monthly Notices of the Royal Astronomical Society 439 (2) (2014) 1245–1269.

[5] S. P. Driver, P. Norberg, I. K. Baldry, S. P. Bamford, A. M. Hopkins, J. Liske, J. Loveday, J. A. Peacock, D. T. Hill, L. S. Kelvin, A. S. G. Robotham, N. J. G. Cross, H. R. Parkinson, M. Prescott, C. J. Conselice, L. Dunne, S. Brough, H. Jones, R. G. Sharp, E. van Kampen, S. Oliver, I. G. Roseboom, J. Bland-Hawthorn, S. M. Croom, S. Ellis, E. Cameron, S. Cole, C. S. Frenk, W. J. Couch, A. W. Graham, R. Proctor, R. De Propris, I. F. Doyle, E. M. Edmondson, R. C. Nichol, D. Thomas, S. A. Eales, M. J. Jarvis, K. Kuijken, O. Lahav, B. F. Madore, M. Seibert, M. J. Meyer, L. Staveley-Smith, S. Phillipps, C. C. Popescu, A. E. Sansom, W. J. Sutherland, R. J. Tuffs, S. J. Warren, GAMA: towards a physical understanding of galaxy formation, Astronomy and Geophysics 50 (5) (2009) 5.12–5.19. `arXiv:0910.5123`, `doi:10.1111/j.1468-4004.2009.50512.x`.

[6] T. Kohonen, The self-organizing map, Neurocomputing 21 (1) (1998) 1–6.

[7] P. Schneider, M. Biehl, B. Hammer, Adaptive relevance matrices in Learning Vector Quantization, Neural Computation 21 (12) (2009) 3532–3561.

[8] M. Biehl, B. Hammer, T. Villmann, Prototype-based models in machine learning, Wiley Interdisciplinary Reviews: Cognitive Science 7 (2) (2016) 92–111.

[9] S. Sreejith, S. Pereverzyev Jr, L. S. Kelvin, F. R. Marleau, M. Haltmeier, J. Ebner, J. Bland-Hawthorn, S. P. Driver, A. W. Graham, B. W. Holwerda, et al., Galaxy And Mass Assembly: automatic morphological classification of galaxies using statistical learning, Monthly Notices of the Royal Astronomical Society 474 (4) (2017) 5232–5258.

[10] L. Breiman, Random forests, Machine Learning 45 (1) (2001) 5–32.

[11] J. Liske, I. Baldry, S. Driver, R. Tuffs, M. Alpaslan, E. Andrae, S. Brough, M. Cluver, M. W. Grootes, M. Gunawardhana, et al., Galaxy And Mass Assembly (GAMA): end of survey report and data release 2, Monthly Notices of the Royal Astronomical Society 452 (2) (2015) 2087–2126.

[12] S. Turner, L. S. Kelvin, I. K. Baldry, P. J. Lisboa, S. N. Longmore, C. A. Collins, B. W. Holwerda, A. M. Hopkins, J. Liske, Reproducible k-means clustering in galaxy feature data from the GAMA survey, Monthly Notices of the Royal Astronomical Society 482 (1) (2018) 126–150.

[13] Y. A. Gordon, M. S. Owers, K. A. Pimbblet, S. M. Croom, M. Alpaslan, I. K. Baldry, S. Brough, M. J. Brown, M. E. Cluver, C. J. Conselice, et al., Galaxy and Mass Assembly (GAMA): active galactic nuclei in pairs of galaxies, Monthly Notices of the Royal Astronomical Society 465 (3) (2016) 2671–2686.

[14] A. Wright, A. Robotham, N. Bourne, S. Driver, L. Dunne, S. Maddox, M. Alpaslan, S. Andrews, A. Bauer, J. Bland-Hawthorn, et al., Galaxy and mass assembly: accurate panchromatic photometry from optical priors using LAMBDAR, Monthly Notices of the Royal Astronomical Society 460 (1) (2016) 765–801.

[15] D. G. York, J. Adelman, J. E. Anderson Jr, S. F. Anderson, J. Annis, N. A. Bahcall, J. Bakken, R. Barkhouser, S. Bastian, E. Berman, et al., The Sloan Digital Sky Survey: Technical summary, The Astronomical Journal 120 (3) (2000) 1579.

[16] A. Edge, W. Sutherland, K. Kuijken, S. Driver, R. McMahon, S. Eales, J. P. Emerson, The VISTA Kilo-degree Infrared Galaxy (VIKING) Survey: Bridging the Gap between Low and High Redshift, The Messenger 154 (2013) 32–34.

[17] E. L. Wright, P. R. Eisenhardt, A. K. Mainzer, M. E. Ressler, R. M. Cutri, T. Jarrett, J. D. Kirkpatrick, D. Padgett, R. S. McMillan, M. Skrutskie, et al., The Wide-field Infrared Survey Explorer (WISE): mission description and initial on-orbit performance, The Astronomical Journal 140 (6) (2010) 1868.

[18] E. Da Cunha, S. Charlot, D. Elbaz, A simple model to interpret the ultraviolet, optical and infrared emission from galaxies, Monthly Notices of the Royal Astronomical Society 388 (4) (2008) 1595–1617.

[19] E. da Cunha, S. Charlot, MAGPHYS in a nutshell, `http://www.iap.fr/magphys/ewExternalFiles/readme.pdf`, accessed: 2018-07-10.

[20] L. S. Kelvin, S. P. Driver, A. S. Robotham, D. T. Hill, M. Alpaslan, I. K. Baldry, S. P. Bamford, J. Bland-Hawthorn, S. Brough, A. W. Graham, et al., Galaxy And Mass Assembly (GAMA): Structural Investigation of Galaxies via Model Analysis, Monthly Notices of the Royal Astronomical Society 421 (2) (2012) 1007–1039.

[21] C. Y. Peng, L. C. Ho, C. D. Impey, H.-W. Rix, Detailed structural decomposition of galaxy images, The Astronomical Journal 124 (1) (2002) 266.

[22] A. Lawrence, S. J. Warren, O. Almaini, A. C. Edge, N. C. Hambly, R. F. Jameson, P. Lucas, M. Casali, A. Adamson, S. Dye, J. P. Emerson, S. Foucaud, P. Hewett, P. Hirst, S. T. Hodgkin, M. J. Irwin, N. Lodieu, R. G. McMahon, C. Simpson, I. Smail, D. Mortlock, M. Folger, The UKIRT Infrared Deep Sky Survey (UKIDSS), Monthly Notices of the Royal Astronomical Society 379 (2007) 1599–1617. `arXiv:astro-ph/0604426`, `doi:10.1111/j.1365-2966.2007.12040.x`.

[23] T. Kohonen, Self-organizing Maps, Springer, 1997.

[24] M. Biehl, B. Hammer, F.-M. Schleif, P. Schneider, T. Villmann, Stationarity of matrix relevance LVQ, in: 2014 International Joint Conference on Neural Networks (IJCNN), IEEE, 2014, pp. 1–8.

[25] M. Biehl, K. Bunte, F.-M. Schleif, P. Schneider, T. Villmann, Large margin linear discriminative visualization by matrix relevance learning, in: The 2012 International Joint Conference on Neural Networks (IJCNN), IEEE, 2012, pp. 1873–1880.

[26] M. Biehl, no-nonsense GMLVQ demo code, `http://www.cs.rug.nl/~biehl/gmlvq`, version 2.3, accessed: 2018-07-01.

[27] L. Breiman, Classification and regression trees, Routledge, 1984.

[28] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, Journal of Machine Learning Research 12 (2011) 2825–2830.

[29] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Random forests in scikit-learn v0.20.0, settings taken from v0.22.0, `http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html#sklearn.ensemble.RandomForestClassifier`,

`http://scikit-learn.org/stable/modules/ensemble.html#random-forests`, accessed: 2018-10-15.

[30] T. Hastie, R. Tibshirani, J. Friedman, The elements of statistical learning, Vol. 1, Springer series in statistics New York, NY, USA:, 2001.

[31] T. Fawcett, An introduction to ROC analysis, Pattern recognition letters 27 (8) (2006) 861–874.

[32] S. Fouad, P. Tino, Adaptive metric Learning Vector Quantization for ordinal classification, Neural Computation 24 (11) (2012) 2825–2851.

[33] F. Tang, P. Tiňo, Ordinal regression based on Learning Vector Quantization, Neural Networks 93 (2017) 76–88.

[34] M. Siudek, K. Małek, A. Pollo, T. Krakowski, A. Iovino, M. Scodeggio, T. Moutard, G. Zamorani, L. Guzzo, B. Garilli, et al., The VIMOS Public Extragalactic Redshift Survey (VIPERS). The complexity of galaxy populations at $0.4 < z < 1.3$ revealed with unsupervised machine-learning algorithms, Astronomy & Astrophysics 617 (A70).

[35] C. Göpfert, L. Pfannschmidt, J. P. Göpfert, B. Hammer, Interpretation of linear classifiers by means of feature relevance bounds, Neurocomputing 298 (2018) 69–79.

[36] I. Jolliffe, Principal Component Analysis, in: M. Lovric (Ed.), International Encyclopedia of Statistical Science, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 1094–1096. `doi:10.1007/978-3-642-04898-2_455`.
URL `https://doi.org/10.1007/978-3-642-04898-2_455`