Cooperative Feature Selection in Personalized Medicine

Dietlind Zühlke[1], Gernoth Grunst[2], and Kerstin Röser[3]

[1] Fraunhofer IAIS, Sankt Augustin, Germany
[2] Fraunhofer FIT, Sankt Augustin, Germany
[3] University Hospital Hamburg-Eppendorf, Germany

Abstract. In this paper we present a cooperative workflow to choose a subset of feature groups from representations of biomedical objects containing a large number of multiple typed features. The choice is done in cooperation of a computer based decision support system and a human expert of the application domain breast cancer research for personalized medicine. The iterative procedure tries to solve the Paradox of Intelligent Selection [1, cf. p. 30], i.e. the necessity to choose a suitable set of features for reasonable modelling without having modelled the actual picture with the whole set of features. The selection module is part of an enabling environment that shall support the insight into so far not yet modelled influence factors of disease processes.

1 Motivation

For personalized medicine a *stratification* of individual patients to groups of patients reacting similar to some applied therapy has to be found. In order to become relevant in the clinical routine it is necessary that these patient groups can be characterized by a suitably small and specific set of diagnostic findings. In recent years the new area of *systems medicine* aims to reveal systematic relations between "-omics" analyses (often used in the stratification research in pharmacology) and particular findings in cell morphology as well as the general information about the patient (two realms of findings used in the current clinical diagnostic routine). Together with findings about tissue and organ properties and anamnestic data the informations represent a holistic view on the patients situation. If this view spans different layers (e.g. the organism, specific organs, tissues and cells) this representation is called a *multi-layer model* of the patient [2]. This model should give the optimal orientation for the individualized therapy suggestion in a stable and sensitive diagnostic procedure.

The task to identify suitable patient groups as well as a suitable set of features defining them is not cognitively manageable by the human domain experts without adequate *information-technological support*. Reasons for this restriction are the case-related mental models of human experts, their mental focus shifting to recently handled patient cases and the complexity and heterogeneity of the multi-layer models describing the patients. Computational systems with the objective to identify case types based on multi-layer models that support the

The Proceedings of ECML/PKDD 2013 Workshop
Scalable Decision Making: Uncertainty, Imperfection, Deliberation (SCALE)
September 23, 2013, Prague, Czech Republic

cognitive abilities of human experts face several challenges. They have to bridge the cognitive gap between the case-related mental models of the pathological experts and the statistical abilities of automatic systems. The latter are – without the suitable incorporation of domain expert knowledge – often overstrained in biological applications by noise or irrelevant but massive variations. In this sense the *computational decision support system* should induce a constructive interaction [4] between the domain experts and the computational algorithms taking into account the limits of judgement of both.

In the following sections we will introduce a workflow to coordinate the actions of a computer based decision support system and a human domain expert to reveal a suitable choice of feature groups for generating an adequate model of relations e.g. in the systems medicine. The approach was developed during a PhD thesis [5]. The methods and their exemplary testing have been developed in a research project for the improvement of orientation of adjuvant breast cancer therapies.

2 The Application Context – Breast Cancer Research Project Exprimage

2.1 Aim and scope of project

The objective of the Exprimage project was to support *adjuvant therapy suggestions* in breast cancer by incorporating information from several biomedical domains. To achieve this goal, information about the therapies performed (chemotherapy, hormone therapy, radiation therapy) would have been necessary to serve as label for a supervised learning task. This information was not provided. Therefore the documented procedure and achieved results can not be seen as a reasonable proof for a superior mathematical form of feature selection. This could not reasonably be evaluated through comparison with recognition results of other forms of feature selection. Its value can nevertheless be judged by domain experts in so far as it managed to condense vague and so far not biologically validated information into potential relevance patterns that can characterize patient groups. The selected motif of features can be affirmed if pertinent biological explanations are found.

2.2 Data sets

The choice of patient cases and their representing data reflects the research interest of the pathologists in the project. The patient cases were *matched pairs* selected from a larger cohort. This means that the pathological expert chose patient cases that had the *same prognosis* according to the current diagnostic process but that showed a *different progress* of the disease, i.e. one patient survived and the other did not. The cognitive support system should help the pathologists to use additional diagnostic means like automatic image analysis, gen expression analysis, blood parameter analysis that could explain the different courses of the patients.

The cohort that was available for our studies in the Exprimage project consisted of 93 patient cases. The investigated patients' resectioned tumour tissue was older than five years – the time interval that is the pertinent clinical frame to evaluate the further perspective of the disease. The clinical data for the patients were collected in the clinical routine during diagnosis and therapy. The patients were categorized by their follow-up-status into three outcomes with the following distribution:

- Follow-up status one (alive): 50 patients
- Follow-up status two (relapse): 7 patients
- Follow-up status three (dead): 36 patients

For all analyses described later, patients with follow-up status two were neglected as there were not enough data samples and the pathologists rejected the combination with follow-up status three. As the information about the *therapy response* (the label suitable for the actual research question) was missing, the *follow-up status* of the patients was chosen as surrogate label which we refer to in classification. A naive or random classification according to the prior distribution of the follow-up status would yield a recognition rate of $58,1\%$ (the prior of class one). The classification given by the prognosis from the current diagnostic process – the *grading* of the patient – has a recognition rate of 61.4% on the given cohort.

In our subproject we focused on supporting the current diagnostic process (clinical data) using information derived from digitized tumour images marked with different histological stains (image data).

Clinical data. The clinical data reflect findings for every patient according to the current state of the art in diagnosis and prognosis. We exploited a subset of ten clinical data that were available for all patient samples. We refrained from imputation of missing data as there were not enough complete samples for a valid statistical estimation of values to be imputed. Most of the incorporated features are related to the microscopic diagnosis describing cell morphology. They are a complementary part of a multi-layer model of the patient's situation with respect to information that is gained from the images on the tissue level.

The pathologists handle some of the diagnostic features in groups according to the tumour properties they describe: the characterisation of the hormone receptors status, the invasion of vessels and the general TNM classification (a characterization that was suggested by the Union for International Cancer Control (`http://www.uicc.org/`) for determining the stage of massive tumours in general). We used this grouping of the data in our analysis. The currently established diagnostic standard – the grading – and the age were used in single feature groups.

Image data. For every patient we had two kinds of stained tissue slice images: structural and functional stains. The starting point for image analysis in Exprimage were raw digitized microscopic images of stained tumour tissue slices.

These images show the tumour and surrounding tissue and thus interactions of the tumour with supporting and nourishing structures that are potential indicators of the prognosis that are not consistently used in current diagnostic schemes. Together with the pathologists, we developed a multi-step automatic image analysis [6] building on a basic characterization of the tissue types. It calculated feature groups representing the hallmarks of cancer [7,8]. We selected two main concepts of tumour description – heterogeneity [9] and distribution patterns [6] – and analysed them under structural or functional perspectives. The derived feature groups are shown in an overview in table 2.

3 Conception of feature group selection in biomedical research

In biomedical research a set of medical concepts $m_1, \ldots, m_M$ is used to characterize the situation of the patients for the considered disease. To handle this characterization within a computational decision support system, each medical concept m_m is represented by (possibly several) groups of features $g_1^{m_m}, \ldots, g_G^{m_m}$ where one feature group g_g consists of (possibly several) features $f_1^{g_g}, \ldots, f_F^{g_g}$ that are adequately comparable using a specific dissimilarity measure d^{g_g}. In this paper we introduce a workflow to identify a subset of feature groups $\phi_1, \ldots, \phi_F$ that allows a good separation of given patient groups $p_1, \ldots, p_P$.

Feature group selection is reasonable from two perspectives: in the mathematical perspective it helps to avoid the curse of dimensionality [10] while in the medical perspective the set of feature groups to be handled should be suitable for diagnosis and prognosis in clinical routine. Especially this last condition requires the feature group set not only to be reasonably small but also to be interpretable by the human experts. Thus the identified feature groups shall be biologically evaluated with respect to their potential relevance for the patient prognosis.

3.1 State of the art in feature group selection for biomedical research

Barillot et al. [11] review different feature selection methods in the area of biomedical research, including plain statistic feature selection (see [12] for a comprehensive overview). In the application context these basic methods leave some open questions, e.g. how many features to choose. Furthermore standard feature selection methods are able to handle heterogeneous data types integrally [11]. In the research for biomarkers on the molecular level often documented domain knowledge is used for feature selection in terms of known networks of components [11], e.g. protein-protein interaction networks. In systems medicine that tries to integrate different layers of diagnostic findings, there is (yet) no schematic knowledge available in data bases.

There is a general consent that it is valuable to exploit any form of reliable knowledge in order to orient machine based selection processes [11, section 6], e.g. prior knowledge about the existence of groups of features. Working at the level of groups of features can induce biological interpretation, reduce the dimension of the statistical problem and increase the stability of the model [11]. Barillot et al. discuss several possibilities to incorporate the domain knowledge with respect to feature groups, in order to detect the importance of groups within given signatures or vice-versa, build signatures from detected important groups.

Furthermore feature groups can be analysed according to their isolated discriminative power by statistical means of correlation with the class partition. On the other hand they can be used in discriminative modelling where the feature groups are used to learn an integral model. According to Barillot et al. group-sparse linear discrimination [11, equation (6.12) p. 190] exploits the groups structure in the features to be coherent with the model structure which directly results in easily interpretable signatures. This method is based on the logistic loss which in turn is based on single differences between numerical feature values.

3.2 Cooperative feature selection

In our application we consider different types of feature groups like numerical, categorical or functional data. Thus we can not apply group-sparse linear discrimination that is only adequate for numerical feature values in our discriminative group modelling. The interpretability of the models is very important in this application. We thus developed a new group modelling algorithm for *automatic contextual feature group selection* that is a descriptive machine learning algorithm based on *modelling class typical representatives (prototypes)*. Before integrating large number of feature groups that is statistically hard to handle into the automatic process for contextual feature groups selection, we performed a pre-selection that applied human expert knowledge.

The human experts in pathology are thinking in case-related contexts. In this respect the pre-selection has to be rendered possible by an automatic feature group ranking that extracts prototypical representatives of the patient case groups is cognitively adequate and supports the human ability to identify potentially relevant feature groups. To test this hypothesis we implemented four selection strategies shown in figure 1.

We start with a short explanation of this algorithm and will postpone the detailed description of the pre-selection strategies.

Contextual feature group selection method. As the technical selection module for discriminative group modelling is linked to corresponding human interpretations, the processing and results have to be transparent for the domain expert. In the group modelling task, we want to determine feature groups that allow the identification and discrimination of significant and potentially relevant patient groups. We developed a method based on Generalized Learning Vector Quantization [16] that models prototypical representatives of the considered classes to achieve a good coupling to the pathological expert' thinking.

To handle the different types of feature groups we extended GLVQ to a dissimilarity adaptation algorithm – Vector based Generalized Learning Vector

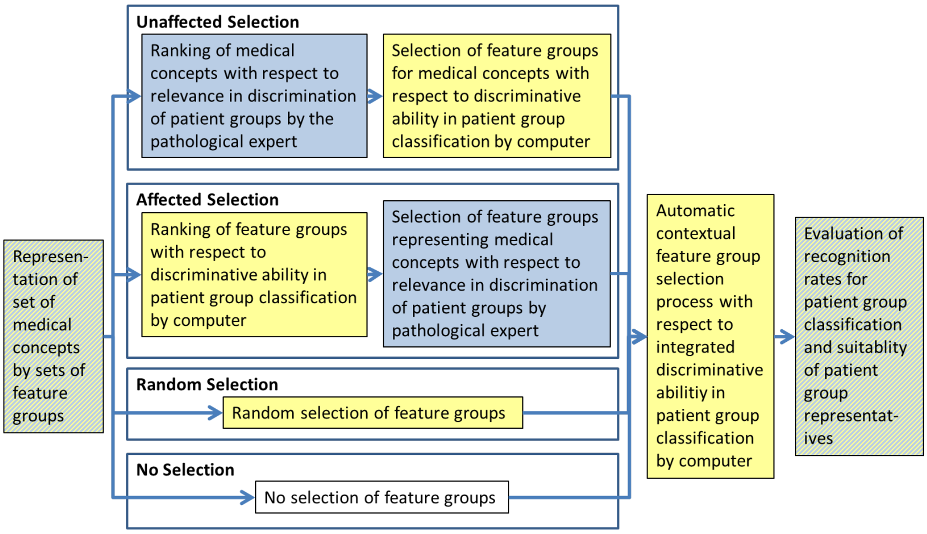
Representation of set of medical concepts by sets of feature groups
Unaffected Selection
Ranking of medical concepts with respect to relevance in discrimination of patient groups by the pathological expert
Selection of feature groups for medical concepts with respect to discriminative ability in patient group classification by computer
Affected Selection
Ranking of feature groups with respect to discriminative ability in patient group classification by computer
Selection of feature groups representing medical concepts with respect to relevance in discrimination of patient groups by pathological expert
Random Selection
Random selection of feature groups
No Selection
No selection of feature groups
Automatic contextual feature group selection process with respect to integrated discriminative abilitiy in patient group classification by computer
Evaluation of recognition rates for patient group classification and suitablity of patient group representat-ives

Fig. 1: Schematic workflow of feature group generation, selection and evaluation. Computer based steps are highlighted in yellow, human processing steps are marked blue. Steps incorporating both are given in grey.

Quantization (vb-GLVQ, [5]) – with specialised dissimilarity measures corresponding to the type of features in a feature group. For comparing two patient cases within this algorithm and to adapt the prototypes accordingly, the dissimilarities d^{g_j} in the single feature groups g_j are combined in a weighted sum. The overall dissimilarity between a patient sample v_k and a prototype w_n is given by:

$$D_\alpha\left(v_k, w_n\right) := \sum_{j=1}^{J} \left(\alpha_j^n\right)^2 d^{g_j} \left(\left[v_k\right]_{[j]}, \left[w_n\right]_{[j]}\right) \tag{1}$$

with $[v_k]_{[j]}$ denoting the feature values in v_k that belong to the feature group g_j and the constraint that $\sum_{j=1}^{J} \left(\alpha_j^n\right)^2 = 1$ for all $n = \{1, \ldots, N\}$ [5]. The weights α_j of the feature groups are adapted within a gradient descent approach that tries to optimize a cost function that represents a maximum margin classifier between different groups of patients. The position of the representatives are adapted in the same manner.

To interpret the dissimilarity parameters identified in this method as relevance values, the dissimilarities within a feature group have to be normalized with respect to their range and variance. In the tests we scaled every dissimilarity within a single feature group by the interquartile range of the pairwise dissimilarities for this feature group in the training data. The interquartile range is the difference between the 75th percentile and the 25th percentile of a sample and is known to be a robust measure of variance, stable against outliers [14]. There are several further possibilities for dissimilarity normalization that we currently analyse for their influence on a suitable automatic selection of feature groups.

The number of free parameters for this method n_{fp} is given by:

$$n_{\mathrm{fp}} = n_{\mathrm{p}} \cdot n_{\mathrm{d}} + n_{\mathrm{f}} \qquad (2)$$

where n_{p} denotes the number of prototypical representatives (one or more per class) and n_{d} is the number of feature dimensions (over all incorporated feature groups) which together form the number of free parameters in the position of the prototypical representatives, while n_{f} denotes the number of incorporated feature groups and accounts for the freedom in the weighting of the dissimilarities in the single feature groups.

Pre-selection of feature groups using cooperative selection strategies According to equation (2) there might be a large number of free parameters to be estimated for a single run of the automatic feature group selection method. In our example that we detail later in section 4 we have 64 feature groups with a total of 175 dimensions, that in the minimum setting of one prototypical representative per class, gives a total number of $2 \cdot 175 + 64 = 414$ free parameters that have to be estimated from 86 patient samples. As this unbalance seems to be statistically challenging we are forced to use a pre-selection of feature groups. We propose different strategies to incorporate the domain knowledge with an automatic evaluation of the discriminative power for the single feature groups.

For the evaluation of the discriminative power of a single feature group, we conducted 20 runs of the vb-GLVQ method introduced in the last section using the single feature group and one prototypical representative per class. For every run, we calculated a new random balanced selection of 30 patient cases for training and 6 patient cases for testing for each class (in total 60 cases for training and 12 cases for testing) and learned for 600 epochs. The discriminative power of a feature group was represented by the average of the test recognition rates that were achieved in the learning task taking into account their variances.

The pre-selection strategies we considered were in detail:

Unaffected selection The pathologists rank the medical concepts represented by feature groups according to the estimated relevance and redundancy. After this medically motivated ranking of the diagnostic features, a prioritization of feature groups is calculated by evaluating the discriminative power of every single feature group. For every concept the chosen number of feature groups that scored highest in the prioritisation is selected for further analysis.

Affected selection The feature groups are automatically ranked according to their single discriminative power. Starting with the best performing feature group, the feature groups are successively selected while skipping redundant diagnostic findings according to the pathological experts.

Random selection To gain a benchmark for the selection process a predefined number of feature groups is randomly drawn from the whole set of feature groups.

No selection A second benchmark is generated by using all available feature groups without any selection.

In the unaffected selection the pathologists judge the relevance of the medical concepts without previous affection or orientation by a computational analysis of the discriminative power of the corresponding feature groups. In the affected selection the results of the discriminative power analysis give a ranking of the feature groups and thus a context of judgement that triggers the selective comment of the pathologists which of the feature groups are actually selected. Both, unaffected and affected selection, incorporate pathological knowledge while neither the random selection nor the incorporation of the whole feature group set (no selection) does.

4 Tests in the application context

The following section describes the different strategies with respect to the achieved recognition results. We will not describe the pathological results of the selection process, e.g. which features were selected. These details can be found in the PhD thesis of Zühlke [5]. Rather we will provide a technical view of the interactive selection process and suggest generalized interaction strategies or schemes.

4.1 Overview of feature groups for selection

In our application example the complete set of 64 feature groups had a total of 175 dimensions. It is given in table 2 at the end of this article. For every feature group we give the full name and the type. We abbreviate numerical descriptors by N. They are handled using the Euclidean distance. The representatives of Gaussian distributions are marked by type G. They are compared using a special type of Kullback-Leibler-Divergence (cf. [5, equation (3.2.6) p. 29]). We handle representatives of distributions with the Cauchy-Schwarz-Divergence (cf. [15] for details of divergence based vector quantization). We abbreviate this type by D. For the relational feature groups we used dissimilarities based on judgements of the pathological experts (RH, cf. [5] for details of their assignment). In the right most column we show the dimensionality of the corresponding feature group that indicates the number of features within the group.

4.2 Test setting

For the feature group sets selected under the different strategies we iteratively applied the automatic contextual feature selection method (both detailed in section 3.2). In every iteration the feature group selection was based on the accumulated weights determined in 20 runs of the contextual method with one prototypical representative per class and learning for 600 epochs. In the balanced setting, we randomly selected 30 patient cases for training and 6 for testing for both classes with different random initialization for every run. In the unbalanced setting, we split the whole patient case set randomly into 72 cases for training and 14 cases for testing in a stratified manner changing random seeds for every run. The cut-off-point for the accumulated weights over the 20 runs was chosen by hand with respect to a significant drop of this relevance measure between two consecutive feature groups. Table 1 shows the best average test recognition rates that were achieved during the iterative process using the preliminary feature group selections according to the different cooperative selection strategies.

4.3 Results for different pre-selection strategies.

In the automatic contextual feature selection based on the set extracted by the *unaffected selection strategy* none of the selected feature group sets achieved a test recognition rate that, taking the variation into account, was significantly better than random or naive classification (random: $51,8\%$). The best recognition rate achieved in one iteration of this selection process was 58.2% with a variation of 8.8% There was no clear pattern in the automatic selection of the feature group sets that showed tendencies of an improvement or decline of the recognition rates. In addition the pathological evaluation of the selection process revealed no comprehensible underlying biological concept. The same holds for the selected feature group sets.

The automatic selection of the feature groups based on the *affected selection* yielded the overall best test recognition rate of 66.7%. Taking into account the standard deviation of 7.6% this test recognition rate was higher than random classification (cf. table 1). In this stage the selection comprised

- four clinical feature groups as well as
- one feature group representing a morphometric clustering of the coarse tumour regions and
- the computationally determined ratio of the expression of the oestrogen receptor.

Further reduction of the affected selection that removed the oestrogen receptor feature group decreased the test recognition rate in all pertinent measures. This indicates that information relevant for the discrimination of the disease courses was dropped. In this case no further reduction of the model complexity and feature group set is possible without a loss of predictive power. The pathologists judged the best performing feature group set to be pathologically interesting and worth further investigation. Figure 2 shows a schematic overview over the automatic selection process using the affected feature group pre-selection.

In the contextual feature group selection process for the *random selection* of feature groups none of the results was better than random classification or the trivial classification according to the classes' prior distribution. The best mean recognition rate was 56.3% with a standard variation of 11.7%.

For the feature group set that was *not selected* the overall second best average test recognition rate was achieved for a selection of two feature groups. With a recognition rate of 65.4% and a variation of 12.2% it was better than random

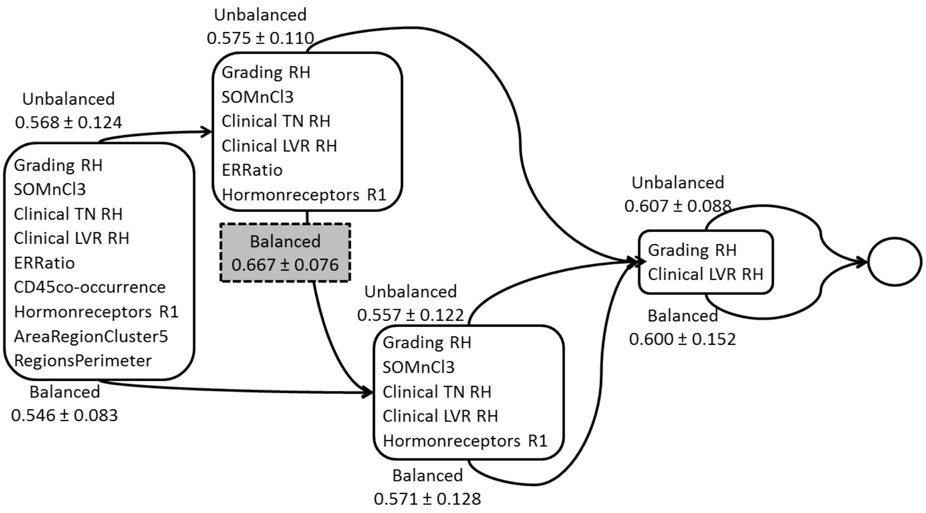

Fig. 2: Schematic overview over the automatic selection process using the affected feature group selection.

classification. However, this combination of two clinical features did not exceed the recognition rate of the current diagnostic process (61.4%) taking its standard variation into account.

The average test recognition rates are shown in an overview in table 1.

Table 1: Best average test recognition rates for the iterative automatic feature group modelling based on different cooperative feature selection strategies

Cooperative selection strategy	Best test recognition rate	Standard variation
Unaffected selection	58.2%	8.8%
Affected selection	66.7%	7.6%
Random selection	56.3%	11.7%
No selection	65.4%	12.2%

4.4 Tentative comparison of recognition rates with other modelling methods

We compared the results of our proposed workflow incorporating the different cooperative pre-selection strategies with a Generalized Learning Vector Quantization [16] using the squared Euclidean distance. While the GLVQ showed very high recognition rates for the training data set (79.1% with standard deviation of 4.7%) the generalization ability given by the test recognition rate (50.4% with standard deviation 9.9%) was significantly lower than for all tests of the cooperative pre-selection and automatic selection workflow using the vb-GLVQ with suitable dissimilarities. That shows that with the high number of free variables estimated from a small set of patient samples there is a tendency to overfit the model to the training data.

For the affected feature group selection we achieved a classification (average test recognition rate 66.7% with a standard variation of 7.6%) that in tendency is better than the classification given by the grading of the patients – the current diagnosis with a recognition rate of 61.4%. The values for the recall and precision of the classes are comparable with the difference that the grading gives slightly higher preference to follow-up status three than our classification.

5 Discussion

We analysed different feature group pre-selection strategies with respect to their suitability to enhance a workflow for feature selection with domain expert knowledge. The quality of the incorporation of domain expert knowledge was judged by the test recognition rates that were achieved using the preliminary feature group selections in an automatic contextual feature selection method as well as by the medical plausibility evaluation of the resulting feature group sets.

Both non-oriented methods of selecting a preliminary feature group set – the *random selection* as well as *no selection* – profit from the automatic contextual feature group selection process. In this process the test recognition rates increased. The whole feature group set was better than random or naive classification with respect to the bias of the classes but not better than the clinical prognosis. The test recognition rate of the random selection did not exceed that of random or naive classification if the standard deviation is taken into account. The random selection did not comprise any of the feature groups identified as relevant in the other tests or by the pathologists.

For the *unaffected selection* the automatic feature selection did not show significant improvements. Possible reasons therefore are:

- The available training data is not representative to derive the free parameters and consequently a selection criterion.
- The normalization of the dissimilarity values in the feature groups is not adequate and therefore the determined weights can not be used as selection criterion.
- The accumulation of the weight values is not adequate.
- The selection of the cut-off in the accumulated weights is not adequate.
- Relevant information was missed.

While the first four reasons are caused by the structure of the feature selection process, the last reason is concerned with the substantial information available for the process. As only in this case the automatic selection with this structure did not improve recognition results, the most probable reason that this feature group selection failed is that important information is left out or missing. This shows that the pathological knowledge if incorporated too early into the selection process can *miss potentially relevant feature groups.*

The affected selection did profit from the automatic contextual feature selection method up to a certain extent. The method has to be monitored in the achieved test recognition rate in order to avoid oversimplification of the model or the loss of relevant feature groups. The best feature group set identified in the automatic contextual feature selection based on the affected selection (cf. section 4.3) was finally evaluated by the pathologists as showing interesting pathological relations that are worth further research. We expect that with a stable data base and relevant labels the resulting feature group selections reveal pathologically relevant information that is able to adjust adjuvant therapies as it was intended by the research project Exprimage.

6 Summary and conclusion

We described a workflow as well as different interaction strategies to identify relevant contextual feature group selections for the discrimination of disease courses in pathological research for personalized medicine. In the prediction of breast cancer follow-up we could show that using the developed learning and evaluation approaches it is possible to identify so far unknown or rather not considered diagnostic dimensions that are worth further experimental medical research.

Summarizing the discussion of the single strategies for pre-selecting feature groups there is evidence that the unaffected selection strategy is less successful than the affected selection strategy. We think that the domain experts need a context for their relevance evaluation that can be provided by the analysis of the single discriminative power of the feature groups. Furthermore the affected feature selection that incorporates the pathological knowledge is more successful than using the whole feature group set in the automatic contextual feature selection method in terms of higher mean test recognition rate and a lower variance for several runs.

The proposed workflow for feature selection is easily extendible for new feature groups. The challenging part for the incorporation of new data is the determination of a suitable dissimilarity measure in the new feature groups. If they are chosen the workflow should be started anew to account for possible cross-relations between the old and the new feature groups.

References

1. G. Myatt and W. Johnson, *Making Sense of Data III: A Practical Guide to Designing Interactive Data Visualizations*. ITPro collection, Wiley, 2011.
2. E. Klipp, W. Liebermeister, C. Wierling, A. Kowald, and H. Lehrach, *Systems biology: a textbook*. Wiley-VCH, 2009.
3. A. W. M. Smeulders, S. Member, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1349–1380, 2000.
4. N. Miyake, "Constructive interaction and the iterative process of understanding," *Cognitive Science*, vol. 10, no. 2, pp. 151–177, 1986.
5. D. Zühlke, *Vector Quantization based Learning Algorithms for Mixed Data Types and their Application in Cognitive Support Systems for Biomedical Research*. PhD thesis, 2012.
6. J. Bornemeier, "Entwicklung von merkmalen zur bestimmung räumlicher ausbreitungsmuster in histopathologischen gewebeschnitten des mammakarzinoms," Master's thesis, Institut für Computervisualistik, Fachbereich Informatik, Universität Koblenz-Landau, 2011.
7. D. Hanahan and R. A. Weinberg, "The Hallmarks of Cancer," *Cell*, vol. 100, pp. 57–70, Jan. 2000.
8. D. Hanahan and R. A. Weinberg, "Hallmarks of cancer: the next generation.," *Cell*, vol. 144, pp. 646–674, Mar. 2011.
9. E. Khabirova, "Image processing descriptors for inner tumor growth patterns," Master's thesis, Bonn-Aachen International Center for Information Technology (B-IT), University of Bonn, 2011.
10. R. Bellman and R. Corporation, *Dynamic programming*. Rand Corporation research study, Princeton University Press, 1957.
11. E. Barillot, L. Calzone, P. Hupe, J.-P. Vert, and A. Zinovyev, *Computational Systems Biology of Cancer*. Chapman & Hall/CRC Mathematical & Computational Biology, London, UK: CRC Press, 2012.
12. I. Guyon, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.
13. T. Kohonen, "Learning vector quantization for pattern recognition," in *Technical Report TKK-F-A601*, 1986.
14. R. R. Wilcox, *Introduction to Robust Estimation and Hypothesis Testing*. Statistical Modeling and Decision Science, Academic Press, 2nd ed., Dec. 2004.
15. T. Villmann and S. Haase, "Divergence-based vector quantization," *Neural Computation*, vol. 23, no. 5, pp. 1343–1392, 2011.
16. A. Sato and K. Yamada, "Generalized learning vector quantization," in *Advances in Neural Information Processing Systems 8*, (Cambridge, MA, USA), pp. 423–429, MIT Press, 1996.

Clinical data		
Feature group full name	Type	Dim
TN characterization of the tumour	RH	7
LVR characterization of the tumour	RH	5
Hormone receptor characterization	RH	8
Age of the patient at surgery	N	1
Grading	RH	3
Quantification of tissues types		
Feature group full name	Type	Dim
Absolute tissue area	N	2
Relative area stroma to overall tumour	N	1
Size variation of tumour regions	G	2
Perimeter variation of tumour regions	G	2
Number of AE1AE3 tumour regions	N	1
Mean area of AE1AE3 tumour regions to tumour area	N	1
Structural heterogeneity characterization		
Feature group full name	Type	Dim
Distribution of inner tumour structure	D	5
Number of regions of different inner tumour structures	N	5
Area distribution for solid homogeneous structures	G	2
Area distribution for half-homogeneous structures	G	2
Area distribution for heterogeneous structures	G	2
Area distribution for sparse heterogeneous structures	G	2
Area distribution for traces of tumour	G	2
Functional heterogeneity characterization		
Feature group full name	Type	Dim
Relative area of functional marker to tumour parenchyma	N	3
CD45 distribution in tissue types	D	4
ER distribution in tissue types	D	4
PR distribution in tissue types	D	4
CD45 co-occurrence with other functional markers	D	3
ER co-occurrence with other functional markers	D	3
PR co-occurrence with other functional markers	D	3
Number of regions	N	3
Area distribution for tumour regions	G	2
Area distribution for ER positive regions	G	2
Area distribution for PR positive regions	G	2
Number of tumour regions covered by hormone receptors	N	2
Spatial distribution of CD45 in tumour regions	D	3
Spatial distribution of ER in tumour regions	D	3

Spatial distribution of PR in tumour regions	D	3

Structural tumour distribution pattern characterization

Feature group full name	Type	Dim
Mean and std dev of edge lengths in Min. Spanning Tree (MST)	G	2
Variation coefficient and min to max for edge lengths in MST	N	2
Average weighted node degree in MST	N	1
Number of nodes in MST	N	1
Randić index in MST	N	1
Distribution of the node degrees in MST	G	2
Variation coefficient and min to max for node degrees in MST	N	2
Mean and std dev of edge lengths in Delaunay Graph (DG)	G	2
Variation coefficient and min to max for edge lengths in DG	N	2
Average weighted node degree in DG	N	1
Number of nodes in DG	N	1
Cyclomatic number in DG	N	1
Randić index in DG	N	1
Distribution of the node degrees in DG	G	2
Variation coefficient and min to max for node degrees in DG	N	2
Morphometry clustering of coarse tumour regions (two clusters)	D	2
Morphometry clustering of coarse tumour regions (three clusters)	D	3
Morphometry clustering of coarse tumour regions (four clusters)	D	4
Morphometry clustering of coarse tumour regions (seven clusters)	D	7
Morphometry clustering of fine tumour regions (two clusters)	D	2
Morphometry clustering of fine tumour regions (three clusters)	D	3
Morphometry clustering of fine tumour regions (four clusters)	D	4

Functional tumour distribution pattern characterization

Feature group full name	Type	Dim
Ratio CD45 to AE1AE3	N	1
Ratio ER to AE1AE3	N	1
Ratio PR to AE1AE3	N	1
Distribution of RCC8 relations for CD45	D	7
Distribution of RCC8 relations for ER	D	7
Distribution of RCC8 relations for PR	D	7
Linear Distance Quantification for CD45	D	2
Linear Distance Quantification for ER	D	2
Linear Distance Quantification for PR	D	2

Table 2: Overview over all feature groups that we considered for the development of a multi-layer model for breast cancer in Exprimage