ACTIVE LEARNING FOR MAGNETIC RESONANCE IMAGE QUALITY ASSESSMENT

A. Liebgott[2], T. Küstner[1,2], S. Gatidis[1], F. Schick[3] and B. Yang[2]

1 Diagnostic and Interventional Radiology, University Hospital of Tübingen, Germany
2 Institute of Signal Processing and System Theory, University of Stuttgart, Germany
3 Section on Experimental Radiology, University Hospital of Tübingen, Germany

1. Motivation

MR image quality evaluation is mainly performed by human observers (HO) to determine the underlying image quality with respect to a certain diagnostic question. HO evaluations are time-demanding and expensive. Furthermore, the lack of a reference image makes this task challenging. In order to support the HO and automize this process, we extend our previous no-reference MR image quality assessment system [1,2] which is based on a machine-learning model observer with an active learning (AL) loop to reduce the amount of needed labeled training data.

2. MR image quality assessment system

The proposed system including the active learning loop is shown in Fig. 1. 3D and 2D multislice MR images are considered as input.

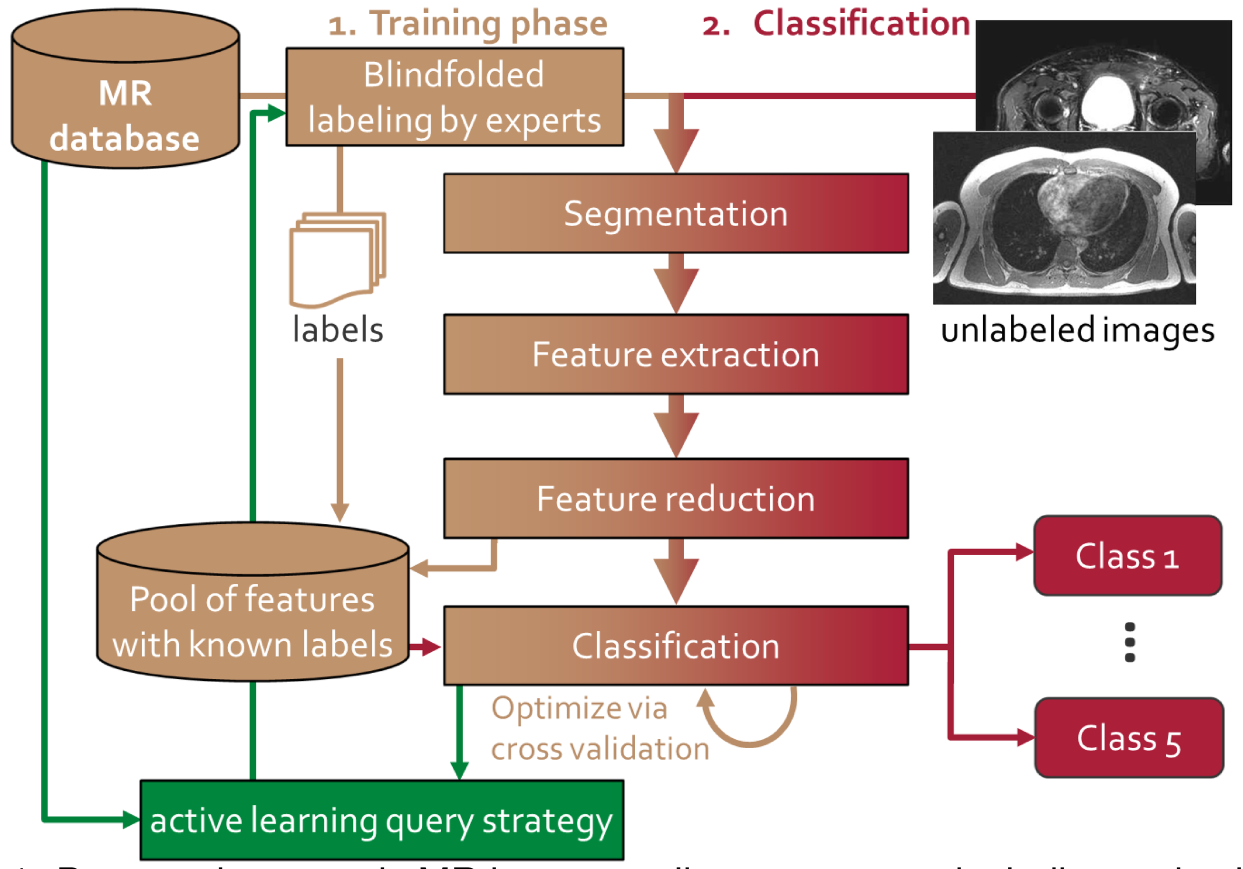


Figure 1: Proposed automatic MR image quality assessment including active learning.

MR database: The database consists of 100 3D images (=2911 2D image slices) from 35 patients of different body regions and imaging sequences.
Labeling: The database is split into 2038 randomly selected 2D image slices/samples for the training set $\mathcal{D}_{\text{train}}$ and 873 samples for the test set $\mathcal{D}_{\text{test}}$. From $\mathcal{D}_{\text{train}}$, randomly drawn samples are assigned to the initial training set $\mathcal{D}_{\text{I}}$ for active learning. All images are blindfolded labeled by experienced radiologists on a 5-point Likert scale, yielding 5 quality classes.
Segmentation: A Chan-Vese segmentation is applied to devide foreground from irrelevant background information.
Feature Extraction: Characteristics such as smoothness, coarseness, regularity, brightness, homogeneity, etc. are applicable interpretations of an MR image, helping to reflect the HO's visual perception. Thus, the MR image is represented by extracted texture-based, intensity-based, gradient-based and fractal-based features, yielding 2871 features.
Feature Reduction: A Principal Component Analysis reduces the dimension of the feature space, yielding a feature vector $\underline{x} \in \mathbb{C}^P, P = 36$.
Classification: Class assignment to class $k \in \{1, \ldots, 5\}$ is accomplished by a soft margin multi-class support vector machine using an one-against-one approach with radial basis function kernel and 10-fold cross validation.

3. Active Learning

The amount of needed labels can be reduced by selecting the most meaningful ones, without redundant information for classification. We implemented two query strategies based on pool-based uncertainty sampling [3], i.e. on how certain the classifier is in his decision. For

$\mathcal{D}$ training data $N_{\text{D}} = |\mathcal{D}|$ number of training samples
$\mathcal{D}_{\text{I}}$ initial training set $N_{\text{I}} = |\mathcal{D}_{\text{I}}|$ number of initial training samples
$\mathcal{L}$ query set $N_{\text{L}} = |\mathcal{L}|$ number of samples per query

the goal is to keep N_{D} as small as possible with an initial training of $\mathcal{D}_{\text{I}}$. In each of the N_q AL loops the HO is queried to label N_{L} samples from the query set $\mathcal{L}$. Furthermore, since HOs label 3D images, but the sample selection takes place on 2D slices, the selection prioritizes samples belonging to the same 3D image.

3.1 probability-based selection

Probability estimates for multi-class SVM derived from pairwise coupling as described in [4] are selected to minimize the difference between the probabilities $P_k(x_n)$ and $P_l(x_n)$ of the 1st and 2nd most probable class [5].

$$\mathcal{L} = \bigcup_{n=1}^{N_{\text{L}}} \{\underline{x}_n | \min_n \left(P_k(\underline{x}_n) - P_l(\underline{x}_n)\right)\}. \quad (1)$$

3.2 distance-based selection

Uncertainty of a sample to belong to class $y_i \in \{1, \ldots, 5\}$ is determined by the distance $d(x_n)$ of one feature vector x_n to the hyperplane $f(x_n)$. The to be labeled set is thus composed by

$$\mathcal{L} = \{\underline{x}_n | d(\underline{x}_n) < d(\underline{x}_m) \; \forall \, x_n \in y_k\} \setminus (\mathcal{O} \cup \mathcal{S}) \quad (2)$$

with the ascending distances and $\mathcal{O}$ denoting a set of outliers and a group $\mathcal{S}$ considering the slack variables (Fig. 2). Distances are determined by

$$d(\underline{x}_n) = \|\underline{w}\|_2^{-1} f(\underline{x}_n) = \|\underline{w}\|_2^{-1} \sum_{i=1}^{N_{\text{SV}}} \alpha_i y_i k(\underline{x}_n, \underline{x}_i) + b \quad (3)$$

where $\underline{w}$ and b denote the primal parameters learned by the SVM with an RBF kernel $k(\underline{x}_n, \underline{x})$ and dual coefficients α_i of N_{SV} support vectors (SV).

Outlier correction $\mathcal{O}$: Reject samples via distance $d_i(\underline{x}_n, \underline{\mu}_i) = \|\underline{x}_n - \underline{\mu}_i\|$ to the class center $\underline{\mu}_i$ of class y_i

$$\begin{aligned}\mathcal{O} = \{&\underline{x}_n | d_i(\underline{x}_n, \underline{\mu}_i) - d_i(\underline{x}_m, \underline{\mu}_i) > \epsilon \\ &\wedge d_i(\underline{x}_m, \underline{\mu}_i) < d_i(\underline{x}_n, \underline{\mu}_i) \\ &\forall \, x_n \in y_i, x_m \neq x_n\}\end{aligned} \quad (4)$$

Slack variable correcton $\mathcal{S}$: Discard samples with minimum distance δ to the hyperplane

$$\mathcal{S} = \{\underline{x}_n | d(\underline{x}_n) < \delta\} \quad (5)$$

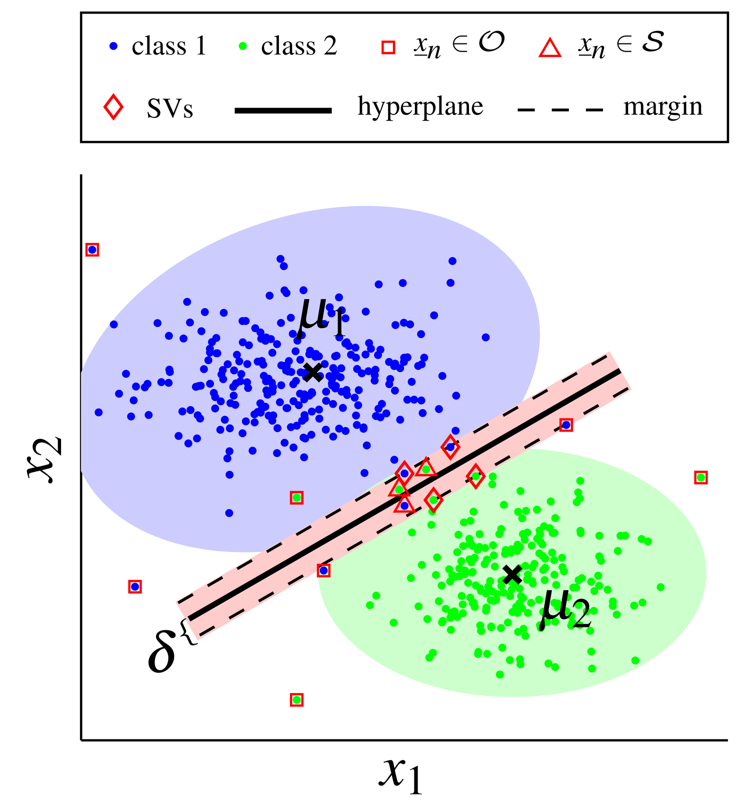
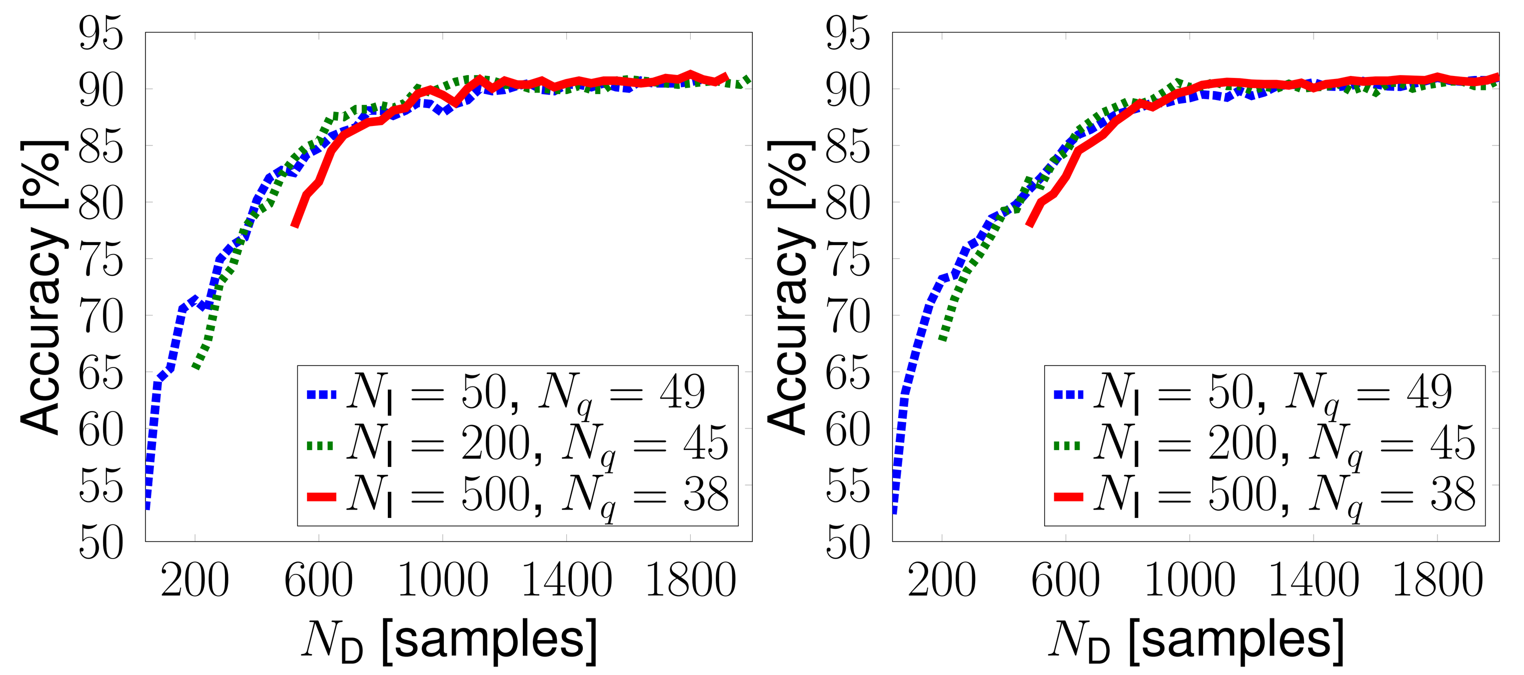


Figure 2: Exemplary 2D feature space to illustrate sample selection.

4. Material and Methods

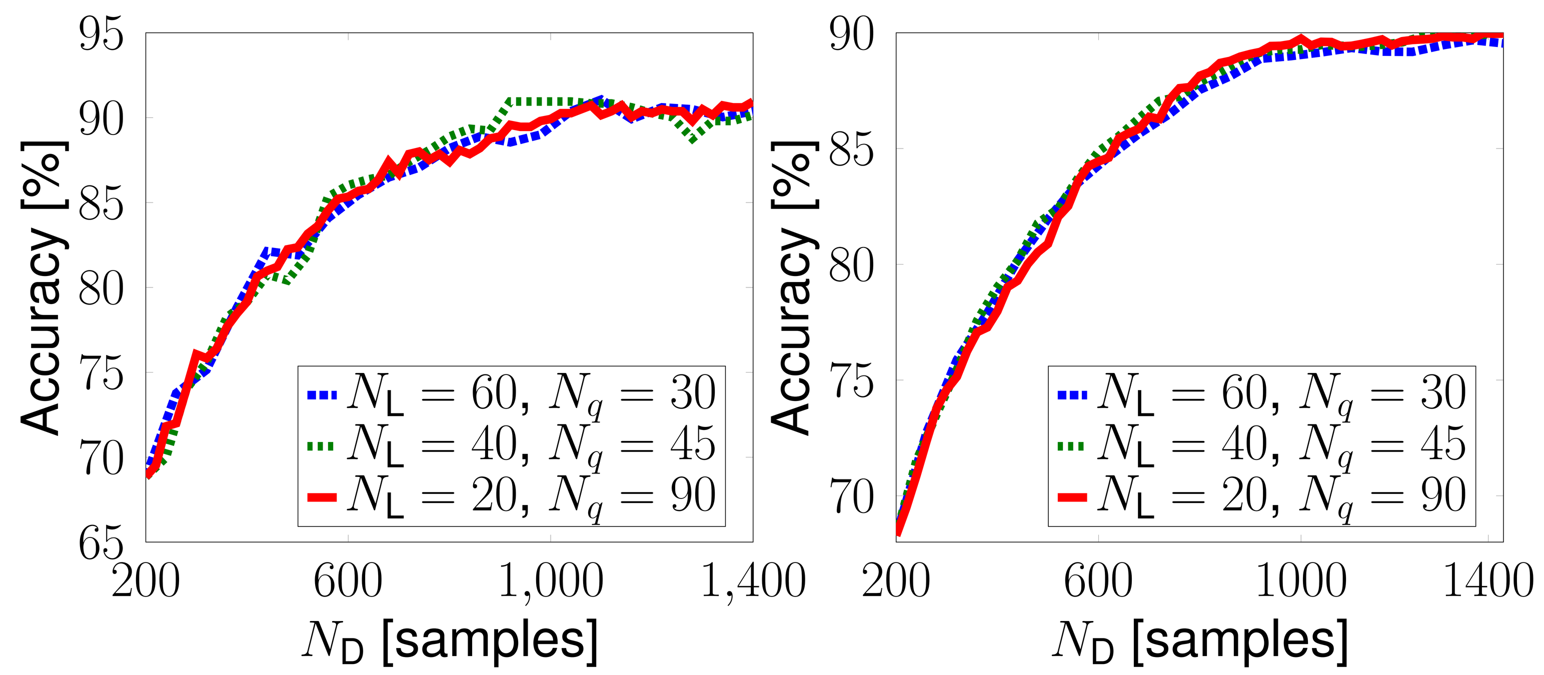
As reported in [1], the system is able to achieve an overall test accuracy of 91.2% with the whole training set $N_{\text{D}} = 2038$. The aim of this study is to reduce the labeling cost, i.e. $N_{\text{D}} = N_{\text{I}} + N_q \cdot N_{\text{L}}$ with N_q queries, while maintaining accuracy. Results are presented as mean of ten different runs.

5. Results



(a) probability-based approach (b) margin-based approach

Figure 4: Test accuracy for initial training set sizes N_{I} with $N_{\text{L}} = 40$ samples per query.



(a) probability-based approach (b) margin-based approach

Figure 5: Test accuracy for N_{L} samples per query with initial training size $N_{\text{I}} = 200$.

6. Conclusion and Outlook

On *in-vivo* MR data both strategies reveal that training data can be reduced by roughly 50% while achieving comparable classification results to the previous system. Furthermore by selecting only the most meaningful 2D slices belonging to a few significant 3D images, the labeling effort is reduced tremendously.

References [1] Küstner et al., *ISMRM Proceedings*, 2015. [2] Küstner et al., *ISMRM Proceedings*, 2016. [3] Lewis and Gale, *ACM Int'l Conf. Res. Dev. Inform. Ret.*, 1994. [4] Wu et al., *J. Mach. Learn. Res.*:5, 2003. [5] Joshi et al., *IEEE Conf. Comp. Vis. Pattern Recogn.*, 2009.

contact information: thomas.kuestner@iss.uni-stuttgart.de

ICASSP 2016, March 20 - March 25, Shanghai, China