Data-Driven Analysis of Dynamic Contrast-Enhanced Magnetic Resonance Imaging Data in Breast Cancer Diagnosis

Thorsten Twellmann

Der Technischen Fakultät der Universität Bielefeld vorgelegt zur Erlangung des akademischen Grades *Doktor der Ingenieurwissenschaften*

Gedruckt auf alterungsbeständigem Papier ISO 9706

Acknowledgements

First of all, I would like express my gratitude to my supervisor Junior-prof. Dr. Tim Wilhelm Nattkemper for his helpful advise and the fruitful discussions during the last years, but also to Prof. Dr. Helge Ritter who aroused my interest in artificial neural networks during my years of study and supported me in various ways throughout the work for this thesis.

Furthermore, I am grateful to Axel Saalbach for the close and productive collaboration and to Dr. Andreas Degenhard for confronting me with such an interesting topic and for initiating the cooperation with the Institute of Cancer Research, UK. I am much obliged to Oliver Lieske for his technical support and Petra Udelhoven for her assistance in handling the manifoldness of bureaucracy.

I warmly thank all members of the Applied Neuroinformatics Group and the Neuroinformatics Group who made M7 to the inspiriting and enjoyable working place that I have appreciated right from the first day. Special thanks go to Tanja Kämpfe, Matthias Kaper, Bert Arnrich and Jörg Ontrup who are also accountable for many amusing evenings.

I thank Matthias Kaper and Volker Wellmann for facing the task of proofreading. I also want to mention the fruitful cooperation with Dr. Peter Meinicke, which result unfortunately has not found its way into this thesis. Christian Müller and Oliver Lichte made their diploma thesis under my supervision and made several important contributions to my work.

Last but not least, I want to thank my family, Kerstin Bruning and my friends for all the support they gave to me.

Publications

Parts of this thesis have been published in advance in:

- IEEE Transactions on Medical Imaging (2005): *An Adaptive Tissue Characterisation Network for Model-Free Visualisation of Dynamic Contrast-Enhanced Magnetic Resonance Data*, T. Twellmann, O. Lichte and T.W.Nattkemper, Note: Accepted.
- BMC Biomedical Engineering OnLine (2004): *Image Fusion for Dynamic Contrast-Enhanced Magnetic Resonance Imaging*, T. Twellmann, A. Saalbach, O. Gerstung, M.O. Leach, T.W. Nattkemper, 3:35.
- Bildverarbeitung für die Medizin (2005): *An Adaptive Colour Scale for Comparison of Pseudo Colouring Techniques for DCE-MRI Data*, T. Twellmann, O. Lichte, A. Saalbach, A. Wismüller, T.W. Nattkemper, Editors: H.-P. Meinzer, H. Handels, A. Horsch and T. Tolxdorff, Springer Verlag.
- Proceedings of the 26th Intern. Conf. of the IEEE Engineering in Medicine and Biology Society (2004): *Detection of Suspicious Lesions in Dynamic Contrast-Enhanced MRI Data*, T. Twellmann, A. Saalbach, C. Müller, T.W. Nattkemper, A. Wismüller
- Proceedings of Medical Imaging (2004): *Visualization of Multivariate Image Data using Image Fusion and Perceptually Optimized Color Scales based on sRGB*, A. Saalbach, T. Twellmann, T.W. Nattkemper, M.J. White, M. Khazen, M.O. Leach, Editor: R.L. Galloway, Jr..

Contents

1 Introduction

In recent decades, imaging became a very powerful tool, offering new possibilities for diagnostic investigations. The development of imaging techniques such as *positron emission tomography* (PET), *computer tomography* (CT) or *magnetic resonance imaging* (MRI) enable physicians to examine pathological disorders in a noninvasive fashion by means of high-resolution three-dimensional images. Recently, the diagnostic power of such methods was further enhanced by the introduction of *multivariate imaging* protocols and setups. In multivariate images, elements of the three-dimensional data array are not only associated with a single signal value but with a *signal vector* of several values. If the images are recorded in a *multimodal* setup, the multivariate nature of the image data arises from the simultaneous application of different imaging techniques (e.g. CT/PET or CT/MRI/PET), whereas in *unimodal* setups, only one technique is utilised for e.g. acquisition of *multitemporal images* consisting of temporal image sequences. Even though multivariate imaging techniques provide new information, integrating and evaluating the much wider range of information is challenging task for human observers. Nevertheless, multivariate imaging techniques have demonstrated to be beneficial in various fields of medical and cognitive sciences and, for instance in breast cancer diagnosis, are moving from research to practical applications.

In the European Union, breast cancer is the most common type of cancer affecting women, responsible for 4% of deaths in the female population [Eustat, 2004] and 17.5% of cancer deaths [Tyczynski et al., 2002]. Breast cancer often affects young women and is the main cause of mortality in women aged between 45 and 64. It is estimated that in the year 2000 there were 350000 new breast cancer cases in Europe. If diagnosed in an early stage, breast cancer has an encouraging cure rate. About 97% of women diagnosed with localised breast cancer will survive five years after diagnosis. Thus, early detection of breast cancer continues to be the key for effective treatment, and screening of the breast with *X-ray mammography* has become part of regular medical check-ups in industrial nations. Recently, *dynamic contrast-enhanced magnetic resonance imaging* (DCE-MRI) has been identified as a valuable complementary technique for breast imaging. In particular for examination of young women, in which X-ray mammography is less helpful due to the denser tissue of the breast, DCE-MRI has demonstrated to be highly sensitive for the detection of cancer. This finding motivated the initiation of several ongoing studies evaluating the potential of DCE-MRI as a screening tool for young women or women at high risk by virtue of genetic predispositions. In DCE-MRI, a temporal sequence of three-dimensional MRI images of the female breast is recorded, depicting the temporal course of the concentration of a contrast agent in breast tissue. The temporal kinetics of the concentration enable radiologists to infer valuable information not only for differentiating between healthy and pathologically affected tissue, but also for distinguishing innocuous benign disorders from life-threatening carcinoma. Even though this information is inherent in the multitemporal image sequence, it does not become evident to the observer by means of the individual images. Hence,

tools are required to aid physicians in analysing the recorded image data.

The aim of the work as presented in this thesis is to develop computational approaches for DCE-MRI data analysis in breast cancer diagnosis. Central component of the presented approaches will be techniques from the field of *artificial neural networks* and *machine learning*[1]. Artificial neural networks allow to analyse DCE-MRI data from a *data-driven* and *model-free* perspective, which differs from the common *model-based* perspective predominant in clinical practice. While in model-based analysis mathematical models describing the measured signal itself or the physiological process underlying the signal are postulated on the basis of existing knowledge, artificial neural networks make possible to derive such models from the DCE-MRI data themselves, eliminating the requirement that physicians enunciate their medical expertise explicitly in a mathematical form.

A central concept of such data-driven approaches to DCE-MRI analysis will be *example-based-learning*: Training signals reflecting the temporal kinetics of contrast agent concentrations in single voxels are exposed to unsupervised artificial neural networks which in turn autonomously reveal categories of similar signals by virtue of their statistical features. Such signal clusters can often be ascribed to different types of tissue such as healthy, benign or malignant tissue and may reveal hidden regularities and structures in the data. Supervised artificial neural networks are able to derive knowledge for the distinction of predetermined classes of signals, e.g. signals of healthy and cancerous tissue, from a sequence of training examples which were assigned by e.g. a human 'supervisor' to one of the considered classes. Therewith, their application is especially expedient if the signal classes can not be modelled explicitly owing to limited a-priori knowledge or high complexity. After *adaptation*, the trained neural network is able to generalise from the seen to unseen examples and can be applied for detecting signals of the corresponding classes in DCE-MRI sequences of new cases.

1.1 Outline of this Thesis

This thesis can be divided into two parts. The imaging technique and concepts of supervised computational learning are presented in the first part. In chapter 2, magnetic resonance imaging and its physical foundations *nuclear magnetism* and *relaxation processes* are briefly described. Various aspects of breast cancer diagnosis by means of dynamic contrast-enhanced magnetic resonance imaging are presented in chapter 3. After a short overview of the anatomy of the breast, pathological disorders and the underlying changes in tissue vascularity are explained. Subsequently, basic principles of DCE-MRI data acquisition and interpretation are presented, including a summary of existing approaches for manual, model-based and model-free data analysis. In chapter 4, two fundamental principals of computational learning are presented. Furthermore, the supervised learning algorithms being utilised in this work and aspects that need to be considered for their application are explained. To the end of the chapter, techniques for comparing the performance of artificial neural networks with e.g. human experts are elucidated.

In the second and central part of the work for this thesis, new applications of artificial neural

[1] Albeit computer scientists performing theoretical research in either field frequently emphasise the distinction of both fields, both terms are used exchangeable in the following due to the thesis' application-oriented focus.

networks in DCE-MRI analysis are introduced, evaluated and discussed. The evaluations are based on DCE-MRI sequences which were provided by the *City Centre Hospital* of the *University of Munich* and by the *Institute of Cancer Research, UK* which is a member of the British *Magnetic Resonance Imaging for Breast Screening* (MARIBS) study. In chapter 5, a system for detection of suspicious lesion masses is presented. Lesion detection is regarded as a binary classification task, and different supervised artificial neural networks are applied to distinguish between healthy and suspicious tissue by means of characteristics of their signals as measured by DCE-MRI. An algorithm which facilities the detailed analysis of the lesion's architecture is introduced in chapter 6. DCE-MRI data of a small group of training cases are correlated by supervised artificial neural networks with label information derived from manual segmentations of the corresponding lesions and their classification in terms of histological examination of biopsy samples. The trained predictors are able to distinguish signals caused by benign, malignant and normal tissue and are applied for augmenting visualisations of DCE-MRI data with pseudo-colours, reflecting the distributions of different signal classes in heterogenous lesion tissue. A technique for systematic comparison of pseudo-colouring methods is presented in chapter 6. Its application is demonstrated by means of a case study, in which the artificial neural networks based pseudo-colouring introduced in chapter 6 is compared with the *three-time-points* method, a model-based clinical standard protocol for DCE-MRI analysis. Visualisation of DCE-MRI data from the perspective of *image fusion* is investigated in chapter 8. Statistical properties of DCE-MRI sequences are evaluated by unsupervised artificial neural networks for combining the individual images of a sequence to new images. The resulting images are compressed visualisations of the entire image sequence and conspicuously display spatial locations of lesion masses. The thesis concludes with a short summary and discussion, followed by a brief sketch of future directions of research.

2 Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) is an imaging technique which has shown to be a valuable source of information in various fields of medical diagnosis including brain diseases, spinal disorder, angiography and tumour diagnosis [Hamm et al., 1994]. The MRI signal arises from the *nuclear magnetism* of hydrogen atoms which are mainly located in the fat and water of the body. By applying an external magnetic field $\mathbf{B}_0$, the spins of nucleons are aligned. Thereby, the nucleons can take over two discrete energetic levels. Excitation by an appropriate radio frequency (RF) pulse causes a change of the spin alignments and absorption of a certain amount of energy proportional to the energetic difference of the two levels. Simultaneously, magnetic field gradients are applied along the three spatial axes to encode the spatial position of the nucleons. After the RF pulse is turned of, the nucleons return to the natural alignment caused by the magnetic field $\mathbf{B}_0$. In doing so, the absorbed energy is emitted as photons which are recorded by the MRI scanner. A three dimensional image can then be obtained by a *inverse Fourier transformation* of the recorded information. Therewith, MRI is a noninvasive and radiation free imaging technique from which three-dimensional images with high tissue contrast can be obtained.

The following chapters give a brief overview of the physical foundation of MRI. A more detailed introduction can be found e.g. in [Webb, 2003] or [Hornak, 2004]. The application of MRI for the acquisition of dynamic contrast-enhanced image sequences for the purpose of breast cancer diagnosis will be discussed in more detail in chapter 3.

2.1 Nuclear Magnetism

The physical effect utilised by MRI is the *spin* of particles such as protons or electrons. A proton spinning around an internal axis has a certain angular moment $\mathbf{P}$ and, as a charged particle, a magnetic moment $\boldsymbol{\mu}$. The relation between the magnitude of both quantities is given by

$$|\boldsymbol{\mu}| = \gamma |\mathbf{P}| \,, \quad |\mathbf{P}| = \frac{h}{2\pi}\frac{\sqrt{3}}{2} \tag{2.1}$$

with the particle-specific *gyromagnetic ratio* γ and *Planck's constant* h. In the following we will focus on hydrogen nucleons with $\gamma = 42.58$MHz/T. Since the magnitude of $\mathbf{P}$ is quantised, so is the magnitude of $\boldsymbol{\mu}$. The magnetic moment can be regarded as a combination of the vectors $(\boldsymbol{\mu}_x, \boldsymbol{\mu}_y, \boldsymbol{\mu}_z)$. In the presence of a strong magnetic field $\mathbf{B}_0$ in the direction of the z-axis, μ_z is given by

$$\boldsymbol{\mu}_z = \gamma \mathbf{P}_z = \gamma \frac{h}{2\pi} m_I \tag{2.2}$$

with the nuclear magnetic quantum number $m_I = \pm\frac{1}{2}$ in the case of a hydrogen nucleon. The difference between the magnitude of (2.2) and the total magnetic moment (2.1) indicates that the

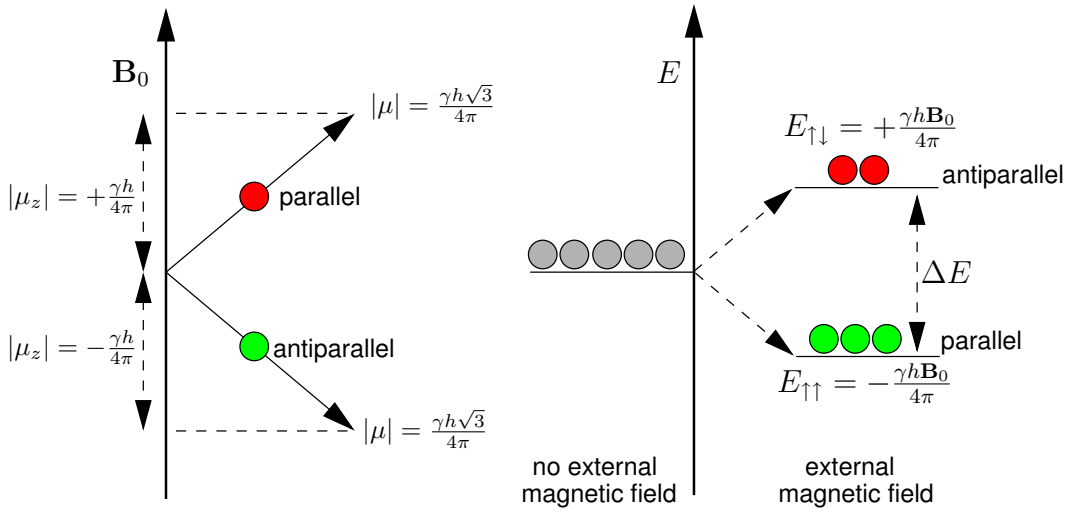


Figure 2.1: Placing a spinning charged particle in an external magnetic field $\mathbf{B}_0$ causes an orientation of the magnetic moment which is parallel or antiparallel to the direction of $\mathbf{B}_0$. Both states correspond to different energy levels $E_{\uparrow\uparrow}$ and $E_{\uparrow\downarrow}$ with an energy difference of $\Delta E = h\nu$ with the *resonance frequency* or *Larmor frequency* ν. In a magnetic field, the high energetic state $E_{\uparrow\uparrow}$ is commonly occupied by a smaller number of particles than the low energetic state $E_{\uparrow\downarrow}$ (Figure adapted from [Webb, 2003]).

magnetic moment is only partially aligned either parallel or antiparallel to the external magnetic field $\mathbf{B}_0$ (Fig. 2.1, left).

Both states can be interpreted as energy levels with

$$E_{\uparrow\uparrow} = -\frac{\gamma h\mathbf{B}_0}{4\pi} \quad \text{and} \quad E_{\uparrow\downarrow} = +\frac{\gamma h\mathbf{B}_0}{4\pi}, \tag{2.3}$$

and the transition of a hydrogen nucleon from the low energy level $E_{\uparrow\uparrow}$ to the high energy level $E_{\uparrow\downarrow}$ can be initiated by a photon with the energy $\Delta E = h\gamma\mathbf{B}_0/2\pi = h\nu$ with ν as the *resonance frequency* or *Larmor frequency*. The same energy is emitted if the transition proceeds in the reverse direction.

If a group of spins is exposed to a magnetic field $\mathbf{B}_0$, each spin is oriented either parallel or antiparallel to $\mathbf{B}_0$ (Fig.2.1, right). The ratio $N_{\uparrow\downarrow}/N_{\uparrow\uparrow}$ between the number of spins in the high energetic state $N_{\uparrow\downarrow}$ and the number of spins in the low energetic state $N_{\uparrow\uparrow}$ depends on the temperature T and is described by the *Boltzmann statistics*

$$\frac{N_{\uparrow\downarrow}}{N_{\uparrow\uparrow}} = \exp\Big(-\frac{\Delta E}{kT}\Big) \tag{2.4}$$

with the Boltzmann's constant k. At room temperature, there are less spins in the high energy state than in the low energy state. Thus, the magnitude of a signal as measured by the MRI scanner is proportional to the population difference $N_{\uparrow\downarrow} - N_{\uparrow\uparrow}$.

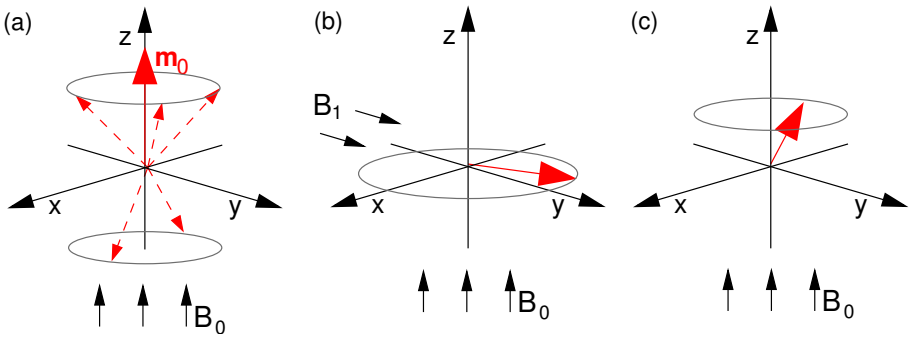


Figure 2.2: (a) In a magnetic field $\mathbf{B}_0$ along the z-axis, the individual spins (dashed) are parallelly or antiparallelly aligned to $\mathbf{B}_0$. The individual spins can be accumulated to a spin-packet with a magnetisation vector along the z-axis. The magnetisation vectors of several spin packets in turn can be accumulated to the net magnetisation vector $\mathbf{m}$. (b) By a second magnetic field $\mathbf{B}_1$ along the y-axis, the net magnetisation can be turned into the xy-plane. (c) After switching $\mathbf{B}_1$ off, the net magnetisation dephases and returns from the xy-plane to the equilibrium (relaxation).

2.1.1 Spin Packets

A group of spins can be regarded as a *spin packet* with a magnetisation vector representing the magnetic field caused by the individual spins. Further on, the magnetisation vector of several spin packages can be accumulated to the net magnetisation vector $\mathbf{m} = (\mathbf{m}_x, \mathbf{m}_y, \mathbf{m}_z)$. At equilibrium, the external magnetic field $\mathbf{B}_0$ and the *equilibrium magnetisation* $\mathbf{m}_0$ are both oriented along the z-axis with $\mathbf{m}_z = \mathbf{m}_0$ (Fig. 2.2, a).

2.1.2 T1/T2 Processes

By exposing the spin packets to photons with a wavelength of the *resonance frequency* or *Larmor frequency* $\nu = \gamma \mathbf{B}_0 / 2\pi$, by e.g. applying a second pulsing magnetic field $\mathbf{B}_1$ along the direction of the y-axis, the net magnetisation vector can be tipped by an angle α towards the xy-plane. For $\alpha = 90°$ the net magnetisation vector rotates in the xy-plane ($\mathbf{m}_z = 0$) with a frequency equal to the Larmor frequency (Fig. 2.2, b). The $\mathbf{B}_1$ pulse needed for an $\alpha = 90°$ tip is called $90°$-*pulse*. From this state, the net magnetisation returns to its equilibrium state (Fig. 2.2, c). This relaxation process is described by

$$\mathbf{m}_z(t) = \mathbf{m}_0 \left[1 - \exp\left(-\frac{t}{\mathrm{T1}} \right) \right] \tag{2.5}$$

and depends on the *spin-lattice relaxation time* (T1).

The second observable effect is the dephasing of the net magnetisation. Immediately after the RF pulse, the individual spins summing-up to the net magnetisation vector are precessing with a coherent phase in the xy-plane. This phase coherence gradually disappears once the pulse is turned off. The return of the transverse net magnetisation $\mathbf{m}_{xy}$ to the equilibrium is described by

$$\mathbf{m}_{xy}(t) = \mathbf{m}_{x_0 y_0} \exp\left(-\frac{t}{\mathrm{T2}} \right) \tag{2.6}$$

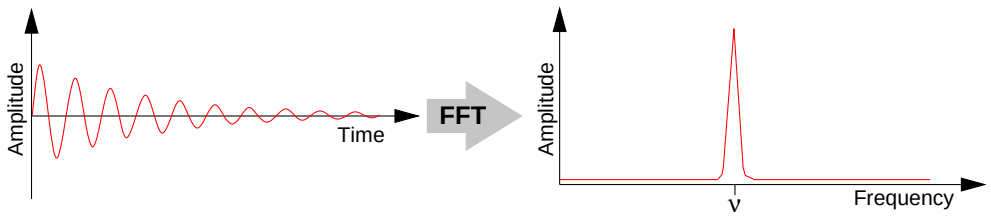


Figure 2.3: The transverse magnetisation rotating about the z-axis can be measured by placing a RF coil around the x-axis. The current induced by the time-varying magnetic flux is a function of time and follows a sine wave that decays due to the dephasing of the spin packets. The frequency domain spectrum of the signal's Fourier transform exhibits a single peak at the *Larmor* frequency.

with T2 as the *spin-spin relaxation time*. The observable time constant $T2^*$ depends on the molecular interactions (described by T2) and on inhomogeneity of $\mathbf{B}_0$ (described by $T2_{\text{Inhomo}}$):

$$T2^* = T2 + T2_{\text{Inhomo}}.$$

Since $T2_{\text{Inhomo}}$ is very small, the observable $T2^*$ is reasonably well approximated by T2 and we will continue to use T2 for the sake of simplicity.

T1 and T2 are tissue-specific constants which are the basis of the tissue contrast of MR images. Thus, valuable information about a tissue segment can be derived by measuring the temporal course of the T1/T2 relaxation process after applying a certain RF-pulse sequence. This is done by measuring the current which is induced during the relaxation process in a RF coil due to the time-varying magnetic flux. For instance, the transverse magnetisation rotating about the z-axis induces a current in a RF coil which is located around the x-axis. Interpreting the current as a function of time yields a sine wave which decays over time due to the dephasing of the spin packets. The measured temporal signal is converted from time- to frequency-domain using the *inverse Fourier Transform*. The frequency spectrum of the Fourier Transform then exposes a peak of certain magnitude at the resonance frequency of hydrogen nucleons in water (Fig. 2.3). Typically, a second peak appears because of the *chemical shift* which causes a slightly different resonance frequency for hydrogen nucleons in fat.

2.2 Magnetic Resonance Imaging

2.2.1 Spatial Decomposition of MR Signals

So far, the measured signal arises from the entire body placed in the magnetic field. Lauterbur realised in 1973 that a spatial variation of the magnetic field causes a resonance frequency that depends on the spatial coordinate of the spin [Lauterbur, 1973]. For instance, a one-dimensional linear gradient $\mathbf{G}_x$ along the x-axis leads to the linearly varying resonance frequency

$$\nu_x = \gamma(\mathbf{B}_0 + x\mathbf{G}_x) = \nu_0 + \gamma x \mathbf{G}_x \tag{2.7}$$

with the magnetic field $\mathbf{B}_0$ and frequency ν_0 at the isocentre. This gradient is also called *frequency encoding gradient* $\mathbf{G}_\nu$.

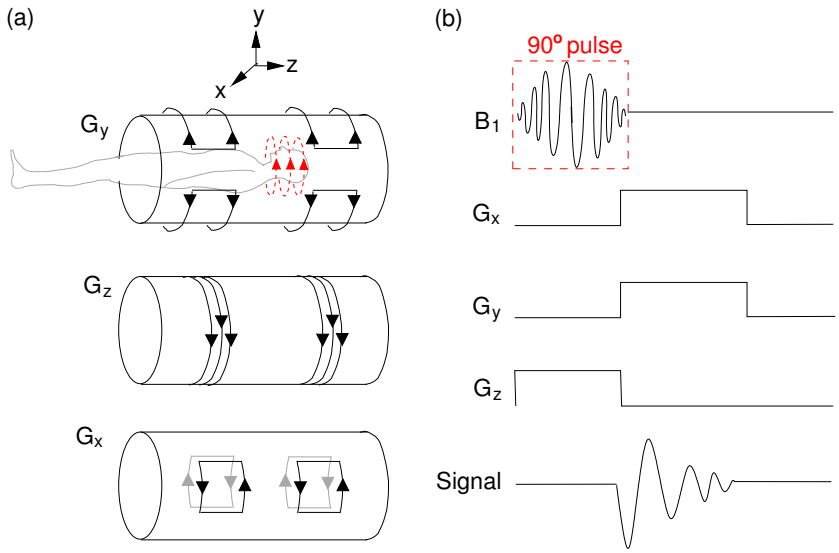


Figure 2.4: (a) The magnetic field gradients along the $x/y/z$-axis are generated by a passage of current through suitably designed coils (black). The RF pulse is generated by an additional coil such as the sketched *bird cage* coil (red) for imaging of the brain. (b) A suitable $90°$-pulse, simultaneously activated with a slice selective gradient (in this case $\mathbf{G}_z$), excites only spins located in a certain slice perpendicular to the direction of the gradient. The spatial resolution in the plane is induced by the frequency encoding gradient $\mathbf{G}_\nu$ which is a linear combination of the gradients along the remaining two axes.

The image information can be recovered by utilising the *backprojection technique* [Lauterbur, 1973], in which NMR spectra are recorded for varying directions of the frequency encoding gradient. For instance, for a yz-plane image of an object, the angle θ between the magnetic field gradient and the y-axis is varied within the range of $[0°; 360°]$. For each angle, an NMR spectrum is recorded which shows several peaks depending on the spatial variation of ν. The sequence of NMR spectra can be backprojected through space by a computer calculating e.g. the *inverse Radon transformation* [Dean and Rodrick, 1983]. The angle θ of the frequency encoding gradient $\mathbf{G}_\nu$ is controlled by linearly combining two gradients $\mathbf{G}_x$ and $\mathbf{G}_y$ according to

$$\begin{aligned} \mathbf{G}_x &= \mathbf{G}_\nu \cos\theta \\ \mathbf{G}_y &= \mathbf{G}_\nu \sin\theta. \end{aligned}$$

A third gradient $\mathbf{G}_{\text{slice}} = \mathbf{G}_z$, in this case along the z-axis, combined with a suitable shape of the pulse with frequency ν_{pulse} determines the z-coordinate of the perpendicular image plane in which spins are measured by affecting only nucleons precessing with a frequency in the range of $\nu_{\text{pulse}} \pm \Delta\nu_{\text{pulse}}$. This gradient is also called *slice selective gradient*. For suitably chosen gradients controlling θ and a suitable $\mathbf{G}_{\text{slice}}$ and ν_{pulse}, the imaging plane can be placed either in *coronal* (perpendicular to the y-axis), *axial* (perpendicular to the z-axis) or *sagittal* (perpendicular to the x-axis) orientation[1].

[1] By convention, the z-axis is in the direction of the head-to-foot axis, the y-axis is along the spine-to-abdomen axis and the x-axis from side-to-side.

It is important to note the temporal order in which the different gradients are applied (Fig.2.4). The slice selective gradient $\mathbf{G}_{\text{slice}}$ is simultaneously applied with the RF pulse and causes only spins in a slice perpendicular to the gradient to be affected by the pulse. After the pulse and the slice selection gradient are turned off, the frequency encoding gradient $\mathbf{G}_\nu$ composed of $\mathbf{G}_x$ and $\mathbf{G}_y$ is turned on and the current induced into the RF coil is recorded.

Rather than using a single pulse for exciting spins in a certain slice, MR images are measured using sophisticated pulse-sequences such as the *spin-echo-sequence*. In this case, a $90°$-pulse is emitted each TR seconds (*repetition time*), followed by an $180°$-pulse after TE seconds (*echo time*) causing the spins to partially rephase and produce an echo signal.

2.2.2 T1/T2-Weighted Imaging Sequences

The pixel intensity $I(x, y)$ of an MR image acquired using a spin-echo sequence is described by

$$I(x,y) \propto \rho(x,y) \underbrace{\left(1 - \exp\left[-\frac{\mathsf{TR}}{\mathsf{T1}}\right]\right)}_{\text{T1-weighting}} \underbrace{\exp\left[-\frac{\mathsf{TE}}{\mathsf{T2}}\right]}_{\text{T2-weighting}}$$

with $\rho(x, y)$ as the proton density. Varying values for TR and TE control the sensitivity of the signal to the T1/T2-relaxation process and result in MR images with different image characteristics. If TR is set to a value much larger than the T1 value of any tissue in the region-of-interest (ROI), the T1-weighting term converges to zero and the sensitivity of the signal to the T1-relaxation process diminishes. The same effect is yielded for the T2-weighting term by choosing TE much smaller than the T2 value of any tissue within the ROI. Cancelling both terms yields the *proton density image* in which the pixel intensity $I(x, y)$ depends solely on the proton density $\rho(x, y)$.

In practice, the TR and TE parameters have to be adjusted by the operator to reasonable values in order to facilitate the distinction of different types of tissue, blood vessels or other anatomical structures in the MR image. The TR and TE values as well as the design of the RF coils significantly influence the quality of the image information.

2.2.3 Multispectral Magnetic Resonance Imaging

In *multispectral magnetic resonance imaging*, a sequence of three-dimensional MRI images from the same ROI is recorded. Each image of the stack is generated with different T1/T2-weighting. Assuming that the images are correctly registered, i.e. they expose no motion artefacts, each spatial position in the displayed ROI is associated with a vector of intensity values, each reflecting the signal intensity as measured with a different parameterisation of the MR scanner. Multispectral images facilitate the discrimination of different tissue types and are used to e.g. distinguish *gray matter*, *white matter*, *cerebrospinal fluid* from pathological structures such as multiple sclerosis or carcinoma in the brain (see e.g. [Holden et al., 1995, Vaidyanathan et al., 1995, Valdes-Cristerna et al., 2004]).

2.2.4 Contrast Agents

In order to enhance the contrast between certain types of tissue, *contrast agents* (CA) are often used to alter relaxation times. After intravenous administration of the contrast agent, the CA

molecules rapidly distribute throughout the bloodstream before they are eliminated through the kidney. During this time, a signal enhancement can be achieved for tissue with increased vascularity or affinity to the CA molecules. Furthermore, examination of the CA concentration course over time allows to infer information about physiological parameters of the local tissue. The technique for acquisition of such temporal kinetic signals reflecting the temporal course of the local CA concentration is called *dynamic contrast-enhanced magnetic resonance imaging* (DCE-MRI) and will be described in more detail in the following chapter.

Paramagnetic contrast agents such as gadolinium based chelate (Gd-DTPA) shorten the T1 value of tissue. Tissue that tends to accumulate CA molecules rapidly is displayed with increased brightness in T1-weighted images, which in turn can be recorded faster due to the shortened T1 value. In contrast to the paramagnetic CA, *superparamagnetic* CA shorten the T2 and T2* time and are mainly accumulated in healthy tissue. They consists of small magnetic particles which possess high magnetic moments and therewith cause inhomogeneities in the local magnetic field. Superparamagnetic CAs are also called *negative contrast agents*. Whereas the tumour intensity remains unaffected, healthy tissue accumulating the CA molecules appears with suppressed brightness in T2-weighted spin-echo sequences.

2.2.5 Multislice and 3D Imaging

A crucial aspect of MR imaging is the time needed for the acquisition of a single MR image. Since the acquisition time predominantly depends on the desired resolution of the in-plane image matrix and the repetition time TR, a trade-off between acquisition time and image resolution exists. In general, the matrix resolution should be as high as possible to minimise the *partial-volume-effect* and to facilitate examination of small anatomical structures. The partial-volume-effect describes the fact that the signal measured for each voxel arises from the entire tissue in a small, three-dimensional cuboid and, therewith, might be caused by a mixture of different tissue types. The smaller the volume element, i.e. the higher the matrix resolution that displays a certain field-of-view, the less likely the signal measured for a voxel is affected by the partial-volume-effect. Constraints on the acquisition time may arise from the investigated region of the body. For instance in imaging the abdomen, the image should be ideally acquired within a single breath-hold (5 to 25 seconds) in order to avoid motion artefacts. Another example are DCE-MRI sequences in which a larger number of images has to be recorded over a time span of several minutes. These temporal constraints come into conflict with the acquisition time of a single high resolution image matrix which can last up to several minutes depending on the hardware, matrix resolution and repetition time.

Several techniques have been proposed to accelerate image acquisition. In a single slice sequence, only the relaxation of spins located in a single plane determined by the slice selective gradient are recorded. Nevertheless, MRI provides the ability to simultaneously record the relaxation of spins in several parallel slices (*multislice imaging*). Instead of affecting only spins with a resonance frequency in the range of $\nu_{\text{pulse}} \pm \Delta\nu_{\text{pulse}}$, spins with a resonance frequency in one of the n non-overlapping frequency bands $[\nu_{\text{pulse}_1} \pm \Delta\nu_{\text{pulse}}], \ldots, [\nu_{\text{pulse}_n} \pm \Delta\nu_{\text{pulse}}]$ are affected in quick succession and the T1 relaxation of the spins in the corresponding n slices is measured simultaneously. The maximum number of slices that can be examined simultaneously is limited by the ratio TR/TE. Additionally, the slices have to be placed with a sufficient spatial distance

to each other in order to avoid partial excitation of the spins in the slices on either side of the currently selected slice. Whereas the multislice imaging technique acquires the image of a three dimensional volume as a sequence of one or several tomographic slices leading to a temporal shift between certain slices, '*true*' three-dimensional images can be recorded using more sophisticated sequences such as the *three-dimensional fast low-angle-shot* (3D FLASH) sequence.

2.3 Summary

MRI is a nonionizing imaging technique which allows for acquisition of three-dimensional images with high spatial resolution and excellent soft-tissue contrast. The MRI signal arises from the *spin-lattice* and *spin-spin* relaxation of hydrogen nuclei in a static magnetic field after excitation by sequences of electromagnetic pulses at the *resonance frequency* of hydrogen. Encoding of spatial information into the signal is accomplished by superimposing three orthogonal magnetic field gradients, resulting in a spatially dependent resonance frequency and phase of hydrogen nuclei. The signal is measured as a current which is induced in a tuned receiver coil due to the time-varying flux caused by the relaxing nuclei. This signal, which likewise is a function of time, is converted by inverse Fourier transform into the spatial domain to produce an image.

The characteristics of MR images can be altered by varying the imaging parameters *echo time* and *repetition time*. This causes the signal to become more sensitive either to the T1 or T2 relaxation process and allows for enhancing the contrast between specific types of soft-tissue. The contrast between different types of healthy tissue as well as between healthy and pathologically altered tissue can be further improved by adminstration of a contrast agent affecting the relaxation times of tissues.

3 Dynamic Contrast-Enhanced MR Imaging in Breast Cancer Diagnosis

In the European Union, breast cancer is the most common type of cancer affecting women [Eustat, 2004]. About 350,000 women were diagnosed with breast cancer in 2000, representing about 26.5% of all female cancers diagnosed [Tyczynski et al., 2002]. Breast cancer risk is strongly correlating to age. About 80% of all cases occur in women over 50 years of age. At the same time, breast cancer is the most common cancer diagnosed in women under 35 years of age. Several risk factors have been identified such as age at first birth, age at menopause or family history of breast cancer. If diagnosed in an early stage, breast cancer has an encouraging cure rate. About 97% of women diagnosed with localised breast cancer will survive five years after diagnosis. Therefore, large screening studies have been initiated to diagnose breast cancer at a very early stage.

Today, *X-ray mammography* (XRM) is the mainstay of screening and diagnosis of breast cancer. Thereby, the three-dimensional breast is typically displayed as two two-dimensional images (one cranio-caudal view and one lateral-oblique view). Even though XRM is the current standard screening technique with a sensitivity between 69% and 90%, it has been criticised for being less helpful in diagnosing cancer in denser tissue of younger women, adjacent to implants or following surgery or radiotherapy [Rankin, 2000]. The specificity of XRM has been reported to be in the range of 10% to 40% and up to 75% of mammographically detected suspicious masses are benign at biopsy [Rankin, 2000]. Additionally, there are concerns about the repeated radiation exposure of XRM, in particular if it is used as a screening tool for younger women.

In this situation, *dynamic contrast-enhanced magnetic resonance imaging* (DCE-MRI) has shown to have impressive potential for detection, diagnosis and monitoring of abnormalities in the breast of younger women [Brown et al., 2000]. In contrast to XRM, the radiation free DCE-MRI technique provides a detailed three-dimensional image of the breast which allows for detecting lesions with high sensitivity. The acquisition of a temporal sequence of images after administration of a contrast agent into the bloodstream also enables physicians to infer information about the state and architecture of lesions. The temporal kinetic signal associated with each voxel allows for distinguishing normal, benign and malignant tissue by means of individual signal characteristics. The sensitivity of DCE-MRI ranges between 93% and 100% [Brown et al., 2000] with a specificity between 37% and 85% [Orel and Schnall, 2001]. The large diversity of reported specificity values is mainly caused by the lack of general accepted standards for imaging protocols and image interpretation. Several studies (e.g. [Brown et al., 2000]) were initiated to examine the value of DCE-MRI as a screening technique for young women or women at high risk by virtue of genetic predispositions. Current goals are to evaluate the effect of DCE-MRI on clinical decisions, the technical performance of the equipment and to determine the most sensitive and specific imaging parameters.

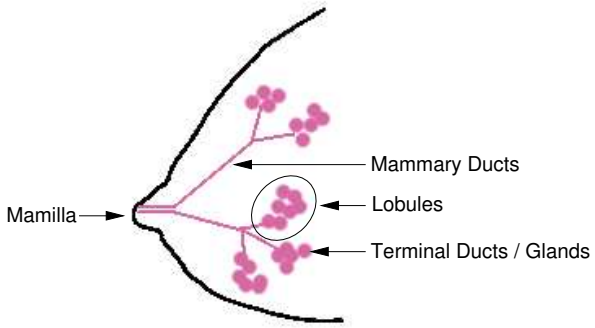


Figure 3.1: Main anatomical structures of the female breast. Malignant disorders mainly appear in the ducts or lobules.

This chapter briefly describes the main aspects of DCE-MRI of the breast for the purpose of breast cancer diagnosis. First, the elementary pathophysiological changes in the tissue of carcinoma are explained, because these are the basis for the detection and characterisation of lesions by DCE-MRI. Starting with the assessment scheme proposed by Kuhl et al., 1999 for the purpose of lesion characterisation by means of visual exploration of temporal kinetic signals, existing model-based and data-driven approaches for *computer aided diagnosis* (CAD) will briefly be reviewed. The interested reader is referred to Heywang-Köbrunner and Beck, 1995 for a more detailed discussion of pathophysiological and radiologic aspects of DCE-MRI based diagnosis of breast cancer.

3.1 Anatomy and Disorders of the Breast

The female breast mainly consists of adipose tissue and connective tissue surrounding the functional structures producing milk during lactation. Starting from the mamilla on the surface of the breast, the functional structure branches to approximatively 20 *mammary ducts* which drain milk from different segments of the breast referred to as *lobes*. Each mammary duct branches to several *lobules* which consist of milk producing *glands*. These glands are connected by the *terminal ducts* with the mammary ducts. In addition to the ducts and lobules, blood vessels and lymph vessels are also especially important for the growth and spread of tumours.

Lesion

Lesion is the general notion for an *abnormal change in structure of an organ or part due to an injury or disease; especially one that is circumscribed and well defined* (Encyclopedia Britannica Online). They are divided into *benign disorders* and *malignant disorders*. Benign disorders are innocuous and do no harm to health whereas malignant disorders tend to infiltrate healthy tissue, to develop metastases, i.e. new lesion masses in other organs, and, in the end, to cause death if not treated effectively. Until a final diagnosis is ultimately determined, each lesion has to be regarded as a suspicious mass which potentially is a malignant disorder.

Benign Disorders

Benign disorders frequently come along with symptoms such as pain or palpable masses which are similar to those of life-threatening carcinoma. The most common are *fibroadenoma* which are solid, smooth, firm lumps. Fibroadenoma are most commonly found in women in their late teens or early 20s but can occur in women of any age. The visual appearance of fibroadenoma depends on the tissue components from which they evolve, but is also influenced by other factors such as age and hormonal status of the patient [Piccoli, 1997].Due to the sometimes similar appearance of certain malignant lesions, fibroadenoma demand for a careful analysis, including various imaging modalities [Heywang-Köbrunner and Beck, 1995]. *Cysts* are fluid filled cavities which also appear as palpable lumps. They are most common in women over the age of 35 who have not reached the menopause. Further types of benign disorders are *nipple disorder*, *mastitis* (inflammation or infection of the breast) or changes caused by open-biopsy, radiation or chemotherapy.

Malignant Disorders or Carcinoma

Malignant disorders or *carcinoma* mainly appear in the ducts and lobules and can be subdivided in *invading* and *non-invading* types. Invading carcinoma such as the *infiltrating ductal or lobular carcinoma*, which are by far the most common type of breast cancer, spread into the surrounding tissue and build metastases. Non-invading types such as the *ductal carcinoma in situ* initially do not break through the walls of ducts or blood vessels and spread only in the duct itself. Nevertheless, if this type of cancer is not recognised early, it may develop the ability to spread into surrounding tissues and, therewith, becomes an infiltrating ductal carcinoma. A detailed assessment of the lesion type and stage can be obtained from a histopathological examination of tissue samples obtained by core needle biopsy.

Vascular Structure of Tumours

Tumour growth commonly comes along with an increased vascular expansion by which the nutrient supply is adjusted to the metabolic needs of cancerous tissue. Two main components of this process are *angiogenesis* and *vasculogenesis*, i.e. sprouting of new capillaries from existing blood vessels and denovo development of blood vessels. Thereby, the vascular structure of cancerous tissue shows several distinguishing characteristics such as spatial heterogeneity of vascular structures, poorly formed and fragile vessels or extreme heterogeneity of vascular density [Collins and Padhani, 2004]. Histopathological examination of tissue samples allow for analysing neovascularity which has shown to be correlated with the tumour grade; a 40% increase in vessel structures doubles the risk of metastases.

In contrast to the histopathological examination of biopsy samples, MRI provides a noninvasive technique for assessing the vascular characteristic of tissue. To this end, the accumulation of *contrast agent* (CA) molecules in the tissue is examined. The temporal progression of this process depends on the delivery of CA (*perfusion*), the ability of CA to leak out of the vasculature (the *vascular permeability*) and the ability of tissue to take up CA (the *extracellular or extravascular volume*) [Leach, 2001]. Since all three properties are affected by the vascular properties of different types of tissue as well as by the pathological changes of vascular structure in cancerous tissue, examination of the temporal kinetic signal describing the temporal course of the CA concentration

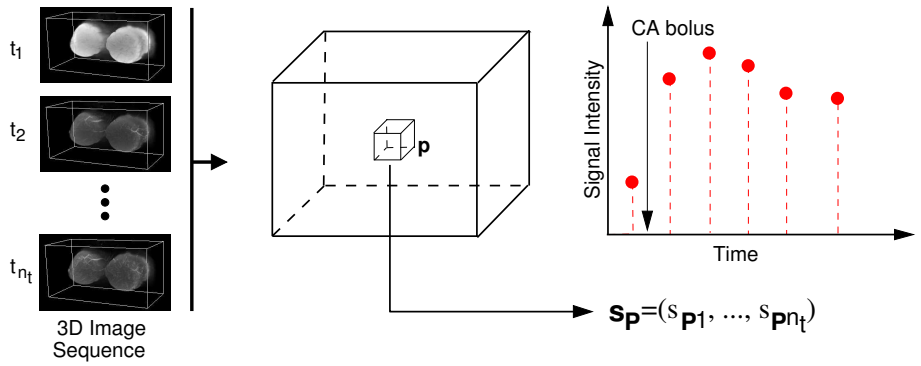


Figure 3.2: The sequence of the three-dimensional pre- and post-contrast images can be interpreted as a four-dimensional image volume. Each voxel with spatial position $\mathbf{p}$ is associated with vector $\mathbf{s_p}$ depicts the temporal course of the signal intensity which reflects the contrast agent concentration in the local tissue.

in the local tissue enables physicians to discern different tissue types as well as different states of lesions.

3.2 Dynamic Contrast-Enhanced MR Image Sequences

For the examination of the process of CA molecule accumulation in tissue, a temporal sequence of n_t MR images is recorded. Although different protocols for DCE-MRI exist, a typical protocol allows for the recording of at least one MR image, referred to as *precontrast image*, before bolus administration of a CA and two or more MR images, referred to as *postcontrast images*, after CA administration. As an example, the DCE-MRI sequence recorded in the *MARIBS* study [Brown et al., 2000] consists of two precontrast and five postcontrast images recorded with a delay of 90sec. Each image of the sequence is a three-dimensional array of *voxels*. The extent of the tissue region represented by a single voxel is frequently anisotrop, i.e. a cuboid with unequal side length.

The signal value $s_{\mathbf{p}t}$ of the voxel with the spatial coordinate $\mathbf{p} = (x, y, z)$ is related to the concentration of CA in the corresponding tissue element at the moment of acquisition of the t-th image. Assuming that the images of a DCE-MRI sequence either exhibit no motion artefacts due to the prone position of the patient in the MR scanner or have been coregistered by a post-processing with a suitable *registration algorithm*, the entire image sequence can be interpreted as a four-dimensional image volume with three spatial and one temporal dimension (Fig. 3.2). The voxel with the spatial position $\mathbf{p}$ is associated with a vector

$$\mathbf{s_p} = (s_{\mathbf{p}1}, \ldots, s_{\mathbf{p}n_t}) \in \mathcal{S} \tag{3.1}$$

describing the temporal course of the CA concentration in the local tissue and can be interpreted as a point in a signal space $\mathcal{S}$.

Depending on the T1/T2-weighting used for the image acquisition, the temporal course of the signal intensity allows to infer information about different physiological parameters of the tissue [Collins and Padhani, 2004]. Using a T2-weighted image sequence, the signal is sensitive to the vascular phase of CA delivery and primarily depends on the physiological properties of *blood*

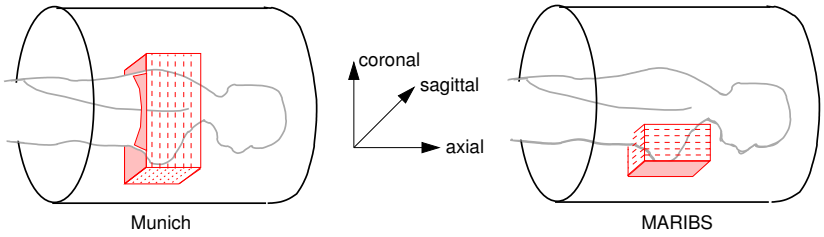


Figure 3.3: Illustration of the extent of the body region employed displayed by the images recorded at the City Centre Hospital of the University of Munich (left) and the MARIBS screening study (right). In both cases, the patient is placed in a prone position. In case of the *Munich* group, both breasts and the major part of the thorax are displayed by the images. The image voxels expose a regular extension in the axial plane. For the *MARIBS* cases, the displayed region of the body is limited to both breasts and the extent of voxels is regular in the coronal plane.

volume and *perfusion*. In a T1-weighted image sequence, the signal is sensitive to the presence of CA in the extravascular-extracellular space (*leakage space*) and thus reflects the physiological parameters *microvessel perfusion*, *permeability* and *extracellular leakage space*. These parameters correlate with the *microvessel density* and the *vascular-endothelial-growth-factor* which induces a characteristic growth of neovasculature [Leach, 2001]. Since the experiments in this thesis only consider T1-weighted image data, the remaining sections will focus on the signal characteristics of T1-weighted image sequences.

3.2.1 DCE-MRI Data Sets

The analytical methods presented in this thesis are evaluated using the DCE-MRI sequences recorded for two different groups of cases. The images of the two groups differ due to application of different imaging protocol, but also in the examined cases. The first group of cases, referred to as *Munich* group, was examined at the City Centre Hospital of the University of Munich and consists of cases, which were all suspicious regarding to a first physical examination and XRM. The second group of cases, referred to as *MARIBS* group, are participants of the *MARIBS* breast screening study which is carried out by several hospitals in the United Kingdom. The study investigates the value of DCE-MRI for the screening of women with a family history of breast cancer suggesting a genetic predisposition.

Munich Data Set

Images of the *Munich* group were acquired with patients placed in a prone position and display both breasts and the major part of the thorax (Fig. 3.3, left). The imaging process was performed with a 1.5 T system[1] equipped with a dedicated surface coil to enable simultaneous imaging of both breasts. First, transversal images were acquired with a STIR (short tau inversion recovery) sequence (TR=$5600ms$, TE=$60ms$, flip-angle=$90°$, TI=$150ms$, matrix size of 256×256 pixels, slice thickness $4mm$), followed by a dynamic T1-weighted gradient echo sequence (3D FLASH, TR=$12ms$, TE=$5ms$, flip angle=$25°$) consisting of 32 to 34 slices in transversal slice orientation

[1] Magneton Vision, Siemens, Erlangen, Germany

Table 3.1: Histological classification and size (number of marked voxels) of the six benign lesions in the *Munich* group.

Case ID	Lesion classification	Size
B1	Fibroadenoma	169
B2	Fibrous mastopathy	497
B3	Scar	26
B4	Lymph node	113
B5	Granuloma with signs of inflammation	99
B6	Chronic mastitis	25

Table 3.2: Histological classification and size (number of marked voxels) of the six malignant lesions in the *Munich* group.

Case ID	Lesion classification	Size
M1	Ductal carcinoma	207
M2	Scirrhous carcinoma	68
M3	DCIS (ductal carcinoma in situ)	49
M4	Status post mastectomy, multilocullar recurrent ductal carcinoma	743
M5	Ductal papillomatosis, transition into papillary carcinoma	169
M6	Ductal carcinoma	284

with a matrix size 256×256 pixels with pixel size of $1.37 \times 1.37mm$ (in-plane resolution) and an effective slice thickness of $4mm$. The dynamic study consisted of $n_t = 6$ measurements with an interval of $110sec$. The first image was acquired prior to a bolus injection of a paramagnetic contrast agent[2], immediately followed by the other five measurements.

The group consists of twelve cases which were suspicious according to a first physical examination and XRM. Subsequent to the acquisition of the DCE-MRI data, the images were interpreted by a radiologist with several years of experience in breast cancer diagnosis. For this purpose, the DCE-MR images were displayed with a clinical standard software. After correlating the DCE-MRI information with the XRM images, the radiologist marked lesion voxels with a cursor on the screening device. According to the histological examinations of the marked lesions, six lesions are classified as benign (Tab. 3.1) and six as malignant disorders (Tab. 3.2).

MARIBS Data Set

The DCE-MRI sequences of the *MARIBS* group were acquired within the scope of the UK multicentre breast cancer screening study *MARIBS*. The primary objective of the study is to test whether MR imaging can be used with equal or better sensitivity and specificity than XRM for the screening of premenopausal women at high risk of developing breast cancer.

The image data of the eight cases considered in this thesis were acquired for a high-sensitivity MRI examination of both breasts, which is routinely performed during the first visit of the patient

[2] Magnevist, Schering, Berlin, Germany

Table 3.3: Lesion classification as denoted in the DCE-MRI diagnosis form sheet and lesion size (number of marked voxels) of the benign cases in the *MARIBS* group.

Case ID	Primary disorder	Secondary disorder	Size
B006A	Fibrocystic change, sclerosing lymphocytic lobulitis	-	324
B013A	Fibroadenoma, fibrocystic change	-	2657
B015A	Fibroadenoma, fibrocystic change, apocrine metaplasia	-	846

Table 3.4: Lesion classification as denoted in the DCE-MRI diagnosis form sheet and lesion size (number of marked voxels) of the malignant cases in the *MARIBS* group.

Case ID	Primary disorder	Secondary disorder	Size
M005A	Non-invasive lobular, invasive ductal/lobular carcinoma, (grade II)	Fibrocystic changes	2634
M007A	Non-invasive ductal, invasice tubular carcinoma, (grade I)	Fibrocystic change, periductal mastitis, sclerosing adenosis, other	175
M009A	Non-invasive lobular ductal, invasive ductal, (grade II)	Fibrocystic changes	902
M094A	Non-invasive ductal, invasive ductal, (grade III)	-	4307
M104A	Invasive, not assessable, (grade III)	-	668

(Visit A). Each DCE-MRI sequence consists of seven three-dimensional MRI volumes, recorded with a separation of 90sec using a standardised protocol consisting of a fast spoiled gradient echo sequence (FLASH) with TR=$12ms$, TE=$5ms$, flip-angle=35° and a field-of-view=$340mm \times 170mm$ in coronal slice orientation. Each voxel of the $256 \times 128 \times 64$ image volume represents a $1.33mm \times 1.33mm \times 2.5mm$ sized tissue volume. Before recording the third volume, a gadolinium-based contrast agent was administered with a bolus injection.

The DCE-MRI image volumes were evaluated by an experienced radiologist, who marked image regions displaying the primary disorder. Secondary disorders were recorded on the diagnosis form sheet but not marked in the image volume. According to the diagnosis form sheet, five cases exhibit malignant (Tab. 3.4) and three cases benign (Tab. 3.3) disorders.

3.3 Interpretation of DCE-MRI Data

According to Heywang-Köbrunner and Beck, 1995, a typical evaluation of a DCE-MRI study consists of

1. Search for or exclusion of an area of enhancement.
2. Analysis of the enhancement characteristics, which are:
   - Presence, speed and amount of enhancement
   - Presence and speed of washout
   - Morphology

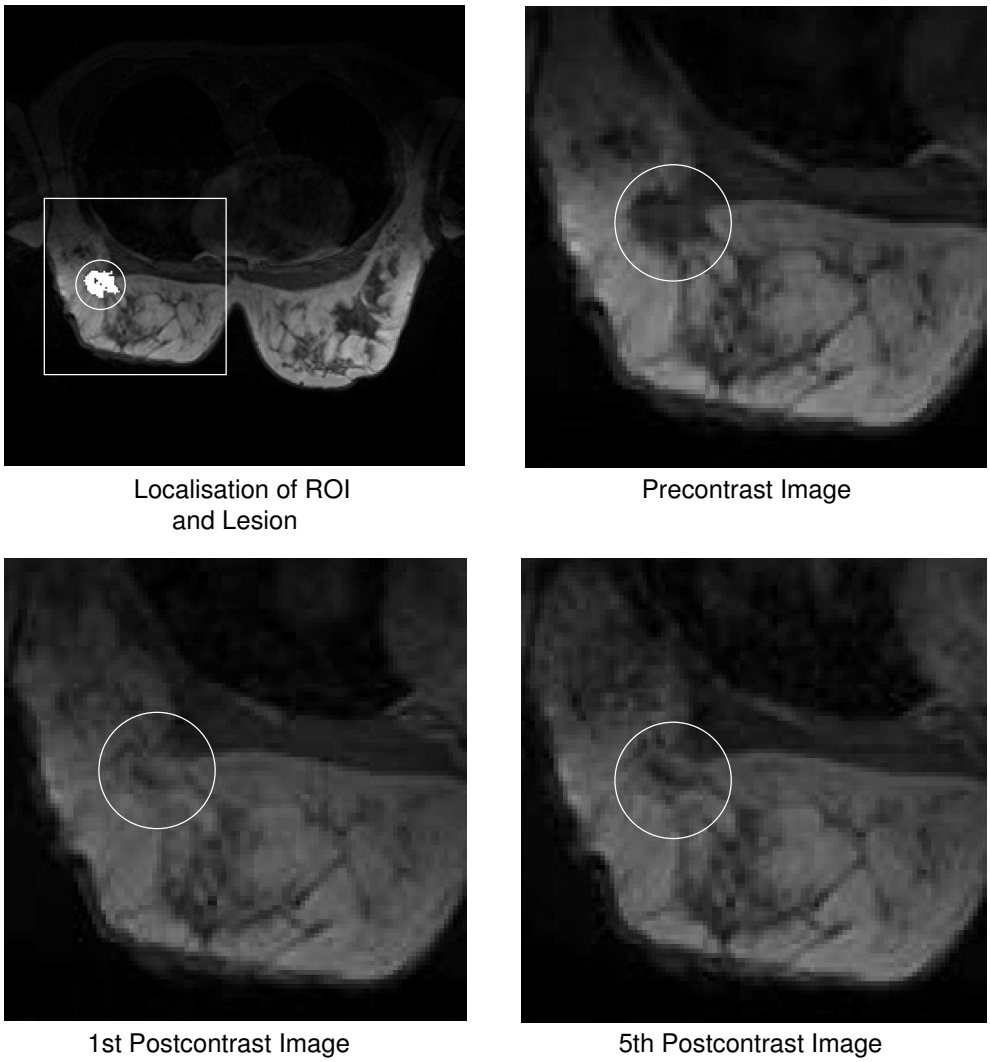


Figure 3.4: Coronal slice of an image volume from a case of the *Munich* group. The upper left image depicts the extent of the ROI (white box) which content is displayed with a zoom factor of 2.5 in the remaining three images. The white circle encloses the lesion whose segmentation is reflected by white pixels. The upper right image depicts the ROI with intensity values reflecting the precontrast signal. The intensity values of the two images in the bottom row reflect signal values in the first and last postcontrast image, respectively. In particular the margin of the lesion exposes a clear signal enhancement in the postcontrast period.

Due to its physiological features, breast tissue affected by benign or malignant disorders tends to accumulate significantly more contrast agent molecules than healthy tissue. Therefore, suspicious tissue masses expose temporal kinetic signals which exhibit signal enhancements in the postcontrast period (Fig. 3.4). A standard approach for detecting such enhancing tissue regions is the examination of *subtraction images*. A subtraction image is computed by subtracting the precontrast image from one of the postcontrast images. The image displays non-enhancing tis-

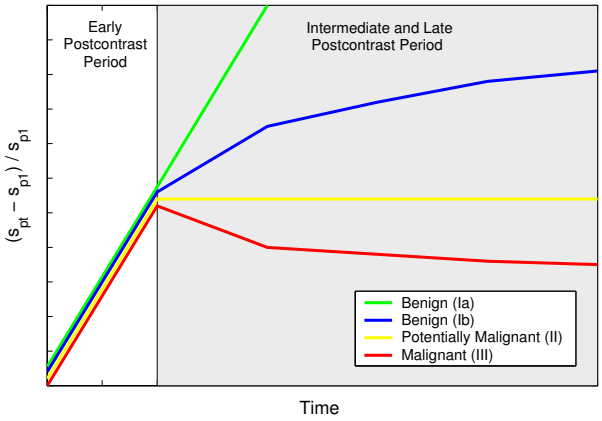


Figure 3.5: Three main types of temporal kinetic patterns can be observed for breast lesions. Signals of type Ia/Ib exhibiting a steady uptake over the entire postcontrast period are likely to be measured for benign tissue. Malignant tissue typically exhibits kinetic signals with a strong uptake followed by a wash-out in the late postcontrast period (Type III). Signals of type II are rated as suspicious for malignancy according to the strong signal uptake followed by a course with indistinct characteristics in the late postcontrast period (Figure adapted from [Kuhl et al., 1999]).

sue with low gray values whereas strong enhancing tissue such as carcinoma appear with high intensities. However, typically a number of tissue regions exists which exhibit a certain signal enhancement, although they are not affected by a pathological disorder. For a thorough evaluation, several subtraction images based on different postcontrast images should be examined. Heywang-Köbrunner and Beck, 1995 propose that at least the subtraction images based on one early and one late postcontrast image have to be considered in addition to the original images of the DCE-MRI sequence.

Subsequent to the localisation and delineation of the extent of lesions, they have to be examined with respect to signs which are indicative for benign or malignant disorders of the local tissue. For this purpose, the radiologist evaluates the enhancement patterns of different lesion voxels and the morphology of the entire tissue mass.

3.3.1 Enhancement Patterns

The temporal kinetic signal $\mathbf{s}$ enables experienced radiologists to infer qualitative information about the physiological parameters of the local tissue. As these physiological parameters are strongly influenced by the type and state of the tissue, distinct temporal kinetic patterns can be observed for different types of tissue. Temporal kinetic signals can either be measured for single voxels or for a larger, manually marked *region-of-interest* (ROI). The examination of signals associated with single voxels provides the most detailed information about the lesion. Nevertheless, the examination of all voxels of a lesion is extremely time-consuming and becomes quickly impracticable for lesions of middle or large size. Further more, the assessments of individual temporal kinetic signals have to be correlated with those of the neighbouring voxels, because isolated signals with a suspicious temporal course are frequently caused by e.g. a larger vessel and not by a disorder of the tissue. Thus, lesions are typically analysed by displaying the average kinetic

signals of all voxels inside manually placed ROIs. Evaluation of entire lesions by means of average kinetic signals of a small number of manually placed ROIs substantially reduces the expenditure of time. However, the ROIs have to be placed very carefully in order to avoid an effect which is comparable to the partial-volume effect: Averaging the kinetic signals of a ROI which covers a larger number of voxels of healthy and cancerous tissue may lead to an indistinct time course signal which misleadingly indicates unsuspicious tissue.

According to Kuhl et al., 1999, three major types of temporal kinetic signals can be distinguished for lesion tissue. The first type of signal exhibits a CA concentration that either continuously increases (Type Ia, Fig. 3.5, green) or flattens in the intermediate and late postcontrast period (Type Ib, Fig. 3.5, blue). The second type of signal (Type II, Fig. 3.5, yellow) shows an initial uptake of CA concentration in the early postcontrast period followed by constant CA concentration in the intermediate and late postcontrast period. Finally, signals of type III (Fig.3.5, red) expose a clear *wash-out* of CA concentration in the intermediate and late postcontrast period after a significant signal uptake in the early postcontrast period. Non-lesion tissue such as fat or muscles typically exhibits temporal kinetic signals which do not show any or only weak enhancements over the entire considered period of time. For the categorisation of the signal time curves, the short-time series are plotted as *relative enhancement curves* as calculated by

$$s_{\mathbf{p}t} = \frac{s_{\mathbf{p}t} - s_{\mathbf{p}1}}{s_{\mathbf{p}t}} \cdot 100,\ t = 1, \ldots, n_t \tag{3.2}$$

with $s_{\mathbf{p}1}$ as the signal intensity at position $\mathbf{p}$ in the precontrast image.

Kuhl et al., 1999 showed that type Ia/Ib signals are likely to be yielded from benign lesions such as *fibroadenoma* (fibrocystic changes). Type III signals are rated as indicative for malignant tissue, whereas type II signals are rated as suggestive for malignancy. The provided experiments considering a cohort of 266 cases indicate a sensitivity of 91% and a specificity of 83% (diagnostic accuracy 86%), if lesions are classified using the described categorisation scheme.

3.3.2 Morphological Patterns

Additional information can be obtained from the morphological structure of the entire lesion. Properties such as characteristics of the margin (smooth, lobulated, irregular or spiculated), internal spatial homogeneity of the lesion (homogeneous, intermediate, heterogenous), presence of peripheral enhancements (*rim enhancement*) or internal non-enhancing septation give rise to a malignant or benign state of lesions [Szabo et al., 2003]. For a reliable assessment of morphological characteristics of lesions, MR images need to have high spatial resolution. On the other hand, the assessment of temporal kinetic signals demands a high temporal resolution, which is achieved at the expense of spatial resolution.

In practice, the optimal diagnostic performance is achieved by the combined assessment of the lesion's morphology and its temporal kinetic signals [Orel, 1999]. The combination of both types of diagnostic criteria is also suggested by the experiments of Szabo et al., 2003, who evaluated the impact of the different diagnostic criteria on the diagnostic performance of a proposed scoring scheme. Table 3.5 shows a brief summary of the major characteristics of benign, suspicious and malignant lesions.

Table 3.5: Summary of morphological and kinetic characteristics of benign, suspicious and malignant lesions. Morphological properties describe spatial patterns of entire lesion masses. The temporal kinetic patterns reflect characteristics of the course of temporal kinetic signals which either refer to single voxels or to ROIs consisting of several voxels.

Lesion type	Morphology	Temporal Kinetic Pattern
Benign	Smooth margin, homogenous tissue	Steadily increasing signal, decreasing slope in the late postcontrast period
Suspicious	Poorly defined margin	Strong uptake, plateau in the late postcontrast period
Malignant	Irregular or spiculated margin, heterogenous tissue, rim enhancement	Strong uptake, washout

3.3.3 Challenges of DCE-MRI Data Interpretation

Due to the lack of standardised protocols for DCE-MR imaging, there are no general guidelines for the interpretation of the image data. The definitions of diagnostic criteria such as *strong signal uptake*, *heterogenous enhancement* or *spiculated margin* are subjective and strongly depend on the experience of the investigator. As a consequence, a certain *inter-* and *intra-observer variance* [Leong et al., 1999, Piccoli, 1997], i.e. varying outcomes from different observers or from repeated assessments by the same observer, has been reported in the domain of DCE-MRI [Orel and Schnall, 2001]. The individual interpretation strategy typically depends on the preferences of the investigator, who has to adapt to the given technical capabilities of the imaging hardware such as the temporal and spatial resolution, but also to the patient population with its prevailing histological features.

There has been a variety of attempts to quantify the characteristics of kinetic signals by using measurements such as the *slope of enhancement*. However, manually chosen parameters such as thresholds, above which certain measurements have to be considered as indicative for malignant or benign tissue, still depend on the investigators experience and may vary with the deployed imaging protocol.

Apart from the varying imaging protocols, the multitemporal nature of the image data turns the interpretation of DCE-MR images into a challenging task for a human investigator. The key-information of the DCE-MRI data, i.e. the temporal kinetic of the signals, is fragmented and distributed over the entire image sequence. To take full advantage of the DCE-MRI data, the investigator has to examine all images of the sequence simultaneously rather than subsequently. For the localisation of enhancing tissue regions, the information of several subtraction images has to be correlated to assure that fast as well as slow enhancing lesions are detected. Following the localisation of suspicious masses, the temporal kinetic signals associated with the lesion voxels have to be examined in order to characterise the lesion as malignant or benign. Due to the heterogenous nature of cancerous tissue, average temporal kinetic signals of *whole-lesion ROIs* commonly exhibit only indistinct signal characteristics. In fact, reliable results are only obtained if the kinetic signals of individual lesion voxels or averaged signals of ROIs, carefully placed at the most enhancing regions, are examined (Fig. 3.6). The placement of such ROIs by means of a conventional display of the three-dimensional image data, i.e. two-dimensional slices of the original image volumes and the subtraction images, strongly depends on the expertise of the radiologist

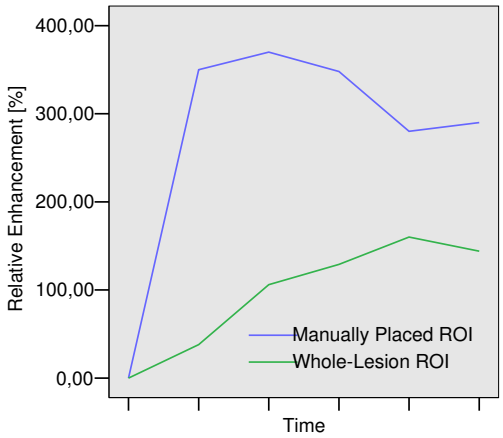


Figure 3.6: Two temporal kinetic signals as measured for a case with a histological proven malignant lesion. The green curve reflects the course of the average signal of a whole-lesion ROI, enclosing all lesion voxels as marked by the manual lesion segmentation. The blue curve depicts the average signal of a ROI which was manually placed over a strong enhancing subregion of the lesion. Due to the heterogeneity of the cancerous tissue, the average signal of the whole-lesion ROI exposes signal characteristics which are more likely for benign tissue. In contrast, the average signal of the manually placed ROI exhibits signal characteristics which are suggestive for malignancy and, therewith, indicates a lesion classification which is concordant with the outcome of the histological examination.

and may vary for repeated examination or for examination by different radiologists.

The continually growing demands for scanners with an increasing spatial and temporal resolution facilitating the examination of ever smaller anatomical structures is the driving force for the rapidly enhancing capabilities of modern imaging hardware. In this situation, the manual evaluation of the multitemporal data by a radiologist may become a limiting factor for the utilisation of such advanced scanners in clinical practice. Radiologists are faced with an increasing amount of information obtained from various, increasingly powerful imaging modalities. As a consequence, there is a substantial demand for CAD systems to assist radiologists in the examination and interpretation of the information by e.g. optimising the display of DCE-MRI data or providing a semiautomatic placement of ROIs to limit the expenditure of time and to attenuate the inter- and intra-observer variability.

3.4 Computer Aided Diagnosis Systems

The acquisition of DCE-MR images as well as the acquisition of images with other imaging techniques are highly computerised processes which lead to an increasing amount of data due to the rapidly enhancing capabilities of modern imaging hardware. To avoid that the human investigator becomes a limiting factor in the diagnostic process, an increasing demand for CAD systems assisting the human expert exists.

Focussing on breast cancer diagnosis based on DCE-MRI data, the application of CAD systems is motivated by the properties of the complex image data, namely:

- A high spatial resolution of MR images displaying the region of interest as a three-dimensional image consisting of several hundreds two-dimensional image slices. These images provide a large amount of spatial information about tissue and anatomical structures. At the same time, the phenomenons under investigation, i.e. pathological disorders of tissue, are typically represented by only a very small fraction of voxels.
- A temporal dimension of data which causes the key-information to be distributed on a large number of three-dimensional images. Subtle differences of certain magnitudes between a very small number of voxels provide evidence to cancerous masses, whereas other signal variations are caused by blood flow or artefacts which mainly stem from respiration and heart beat motions.
- A multiparameter component, which can be introduced by acquisition of images with different T1/T2-weightings. Variation of the influence of the two relaxation processes leads to a varying contrast between different anatomical structures and may facilitate the elimination of tissue regions misleadingly identified as suspicious.

From these properties and from the conceptional formulation of DCE-MRI data analysis [Heywang-Köbrunner and Beck, 1995], the following three main areas of application of CAD systems in DCE-MRI analysis can be formulated:

I **Efficient visualisation of the entire image data** in order to provide access to the key-information of the DCE-MRI data and to facilitate manual data exploration.

II **Localisation of suspicious masses** in order to reduce the amount of image information that has to be evaluated by the investigator by guiding the investigator's attention directly to the spatial locations of suspicious masses.

III **Characterisation of suspicious masses** in order to assist manual examination or to perform a fully automatic classification of lesion compartments or entire lesions resulting in pathophysiological assessments.

In recent years, different CAD systems for DCE-MRI data have been proposed. The area of application of such tools ranges from software providing a platform for data exploration and manipulation (e.g. [Subramaninan et al., 2004], [Engelmeier et al., 2004]) to analytical methods evaluating the measured signal in order to provide an assessment of the local tissue or quantitative information about its physiological parameters. Such analytical methods can be subdivided into two groups. The first group are *model-based* techniques which employ an explicitly formulated mathematical model of the physiological process underlying the recorded signal. The second group consists of *data-driven* techniques which derive implicit models by a data-driven adaptation to the measured image data.

3.5 Model-Based Image Analysis

Model-based approaches such as *pharmacokinetic models* or the *three-time-points* (3TP) method are based on explicitly defined mathematical models describing the physiological process underlying

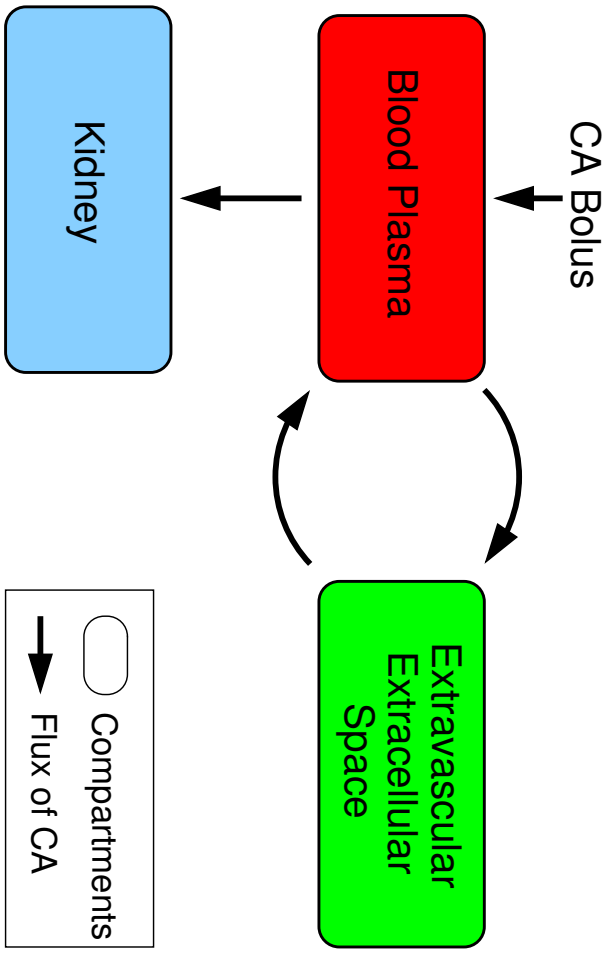


Figure 3.7: Pharmacokinetic models provide mathematical models for the temporal kinetic signals examined in DCE-MR image sequences. The models consider contrast agent concentration in as well as the exchange of contrast agent molecules between tissue compartments. The temporal kinetic patterns as measured by a T1-weighted DCE-MRI protocol reflect the temporal course of contrast agent concentration in the extravascular-extracellular-space compartment.

the measured signal or the course of the measured signal itself. Thus, both methods depend on an extensive a-priori knowledge about the phenomenon under investigation. Evaluation of the measured temporal kinetic patterns with one of these models allows to infer quantitative information about physiological parameters describing the vascular properties of the local tissue.

3.5.1 Pharmacokinetic Models

Pharmacokinetic models [Tofts and Kermode, 1991, Tofts, 1997, Tofts et al., 1999] are widely used for modelling dispersion profiles of molecules such as drugs or contrast agents in the tissue. Physiologically meaningful parameters are combined in a mathematical model describing the temporal change of molecule concentration in the considered tissue compartments (Fig.3.7). Fitting these

models to a recorded signal such as the DCE-MRI signal describing the change of CA concentration enable physicians to infer quantitative information about the underlying physiologic process. This information in turn can be displayed as a *parametric map*. Parametric maps illustrate spatial variations of physiological parameters by voxels displayed in pseudo-colours reflecting the local parameter values.

DCE-MRI considers the dispersion of contrast agent molecules in the tissue. More precisely, the signal of a T1-weighted imaging sequence is predominately sensitive to the quantity of CA molecules in the extravascular-extracellular-space of tissue. The complexity of the physics behind MRI prohibits a complete and detailed model and certain simplifications are necessary. The most frequently used pharmacokinetic models (see [Tofts et al., 1999] for a brief overview) consider only a limited number of compartments such as the blood plasma compartment and the extravascular-extracellular-space compartment. Furthermore, most models assume [Tofts, 1997]:

1. A uniformly distributed CA inside the compartment C_i, C_j.

2. A linear intercompartment flux k_{ij} of CA molecules between two compartments C_i, C_j, i.e. a flux which is proportional to the difference between the CA concentration in both compartments.

3. Time invariant and constant parameters describing the compartments during the period of data acquisition.

The mathematical description of all kinetic models used for DCE-MRI analysis can be ascribed to the *generalised kinetic model* [Tofts et al., 1999]

$$\frac{\partial C_l}{\partial t} = K^{\text{trans}}(C_p - C_l/v_{\text{e}}) = K^{\text{trans}}C_p - k_{\text{ep}}C_l$$

with C_l as the concentration of CA in the lesion tissue and C_p being the concentration of CA in the arterial blood plasma. The parameters of interest are the transfer constant K^{trans} and $k_{\text{ep}} = K^{\text{trans}}/v_{\text{e}}$ with v_{e} being the fractional volume of extravascular-extracellular-space. K^{trans} reflects the flux of CA molecules through the vascular endothelium whereas k_{ep} relates to the flux between the extravascular-extracellular-space compartment and the blood plasma compartment. The combination of both parameters provides information about perfusion and vascular permeability and allows for assessing the 'leakiness' of vasculature, which reflects the angiogenesis within a tumour [Leach, 2001].

In practice, values for K^{trans} and k_{ep} are estimated by a least-square fit of the model to the temporal kinetic signal $\mathbf{s_p}$ as measured at the spatial position $\mathbf{p}$. A voxel-by-voxel evaluation of the entire lesion using the pharmacokinetic model leads to position dependent tuples $(K^{\text{trans}},k_{\text{ep}})_{\mathbf{p}}$. These can be displayed as pseudo-colours superimposed on a two-dimensional (Fig. 3.8) or three-dimensional visualisation of a MR image (e.g. [Hellwig et al., 2002]) and provide a physiologically meaningful visualisation of the data depicting the spatial variation of tumour vasculature throughout the lesion.

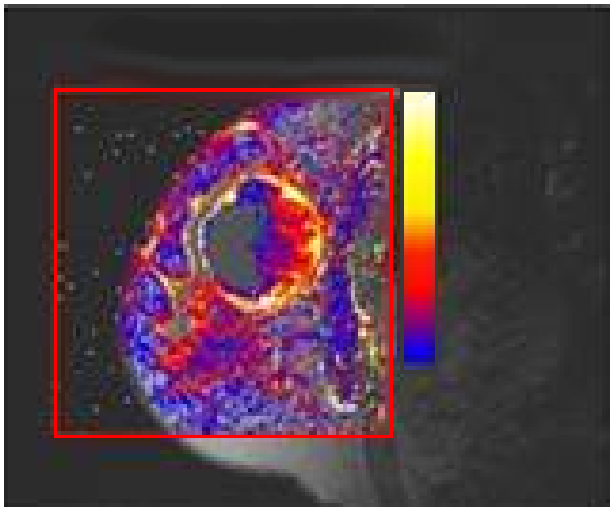


Figure 3.8: Parametric map depicting the local value of K^{trans} as pseudo-colours. The image indicates a high microvessel density (high K^{Trans}) at the margin of the lesion surrounding a necrotic core. Both properties are indicative for malignancy. (Image provided by David Collins, Cancer Research, UK.)

3.5.2 Three-Time-Points Method

Even though the *three-time-points* (3TP) method [Kelcz et al., 2001, Weinstein et al., 1999] is literally not a pharmacokinetic model, it provides a pseudo-colouring of lesion voxels which has shown to be related to the tissue parameters K^{trans} and extravascular-extracellular-space fractional volume v_{e}. A 3TP based voxel-by-voxel colouring of lesions requires DCE-MRI sequences consisting of one precontrast, one early postcontrast and one late postcontrast image as measured at time points t_1, t_2 and t_3, respectively. The strength of the signal uptake in the early postcontrast period is mapped to the intensity of the pseudo colour, whereas the presence of a wash-out is mapped to the colour hue. Both values are obtained from a computational inexpensive model, which allows for a rapid evaluation of lesions.

Signal Model of 3TP

Let each voxel with spatial coordinate $\mathbf{p}$ be associated with a temporal signal triple $\mathbf{s_p} = (s_{\mathbf{p}t_1}, s_{\mathbf{p}t_2}, s_{\mathbf{p}t_3})$. Then, each voxel is mapped to a pseudo colour $\mathbf{c_p} = (h_{\mathbf{p}}, i_{\mathbf{p}})$ with colour hue $h_{\mathbf{p}}$ and intensity $i_{\mathbf{p}}$. The colour intensity is calculated by

$$i_{\mathbf{p}} = \frac{s_{\mathbf{p}t_2} - s_{\mathbf{p}t_1}}{t_2 - t_1}$$

and displays the strength of the signal uptake between the precontrast and the early postcontrast image. For the purpose of visualisation, the intensity corresponding to the strongest signal uptake observed for a certain lesion is scaled to 255. The colour hue reflects presence or absence of a

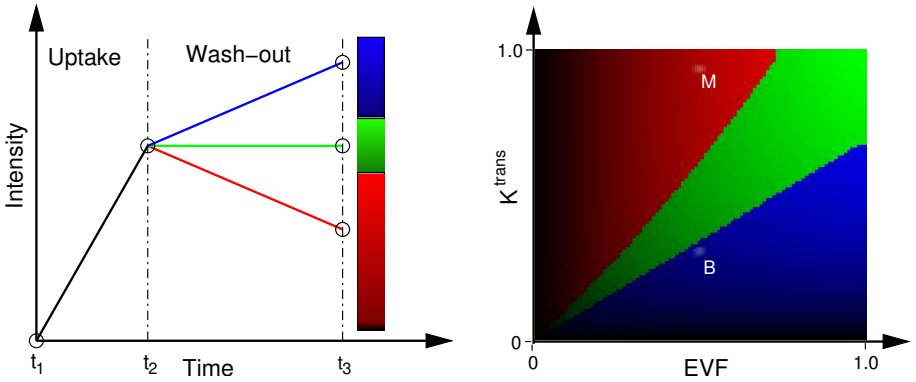


Figure 3.9: Left: Illustration of the 3TP pseudo-colouring scheme. The intensity of the pseudo-colour reflects the amount of signal uptake between the precontrast and the early postcontrast image. The colour hue reflects presence or absence of a significant wash-out in the postcontrast period. Right: The 3TP calibration map illustrates the relation between pseudo-colours and values of the (K^{trans}, v_e) tuple. A typical parameter tuple for malignant (M) and benign (B) tissue is indicated by the white crosses, respectively.

significant signal wash-out between the two postcontrast images:

$$h_{\mathbf{p}} = \begin{cases} \text{red} & : \text{ if } s_{\mathbf{p}t_3} < s_{\mathbf{p}t_2} \ \wedge \ |s_{\mathbf{p}t_2} - s_{\mathbf{p}t_3}| > \sigma s_{\mathbf{p}t_2} \\ \text{blue} & : \text{ if } s_{\mathbf{p}t_3} > s_{\mathbf{p}t_2} \ \wedge \ |s_{\mathbf{p}t_2} - s_{\mathbf{p}t_3}| > \sigma s_{\mathbf{p}t_2} \\ \text{green} & : \text{ else} \end{cases}$$

The parameter σ controls the tolerance for the comparison of the two postcontrast values. Only a signal change of a certain magnitude, e.g. 10% of the early postcontrast value, is rated to be indicative for presence or absence of a significant signal wash-out (see Fig.3.9, left).

Pathophysiological Interpretation and Model Calibration

The pathophysiological interpretation of the 3TP outcome is given by a calibration map which illustrates the relation between pseudo-colours $\mathbf{c}$ and tuples of physiological parameters (K^{trans}, v_e) [Weinstein et al., 1999]. Artificial temporal kinetic patterns are generated for tuples (K^{trans}, v_e) with $K^{\text{trans}}, v_e \in [0; 1]$ using the pharmacokinetic model of Tofts and Kermode, 1991 and are mapped to pseudo-colours by 3TP. These pseudo-colours are displayed at the corresponding positions (K^{trans}, v_e) leading to a two-dimensional calibration map (Fig. 3.9, right). The (K^{trans}, v_e) parameter space is subdivided by 3TP into three regions. Tissue with high microvessel density and permeability (high K^{trans}) and high cell density (low v_e), which in combination is indicative for cancerous tissue, is displayed intense red. A low cell density (high v_e) and a low microvessel density and permeability (low K^{trans}) is indicative for benign tissue and is displayed blue. Both regions are separated by a green region marking parameter tuples with indistinct pathological interpretations.

The relation between (K^{trans}, v_e) tuples and associated pseudo-colours depends on the tolerance parameter σ and the design of the MRI protocol (time points of image acquisition, T1/T2 weighting, etc.). All parameters have to be faithfully selected in order to obtain an optimal pixel-mapping function. In practice, the time points of image acquisition as well as σ are chosen for a

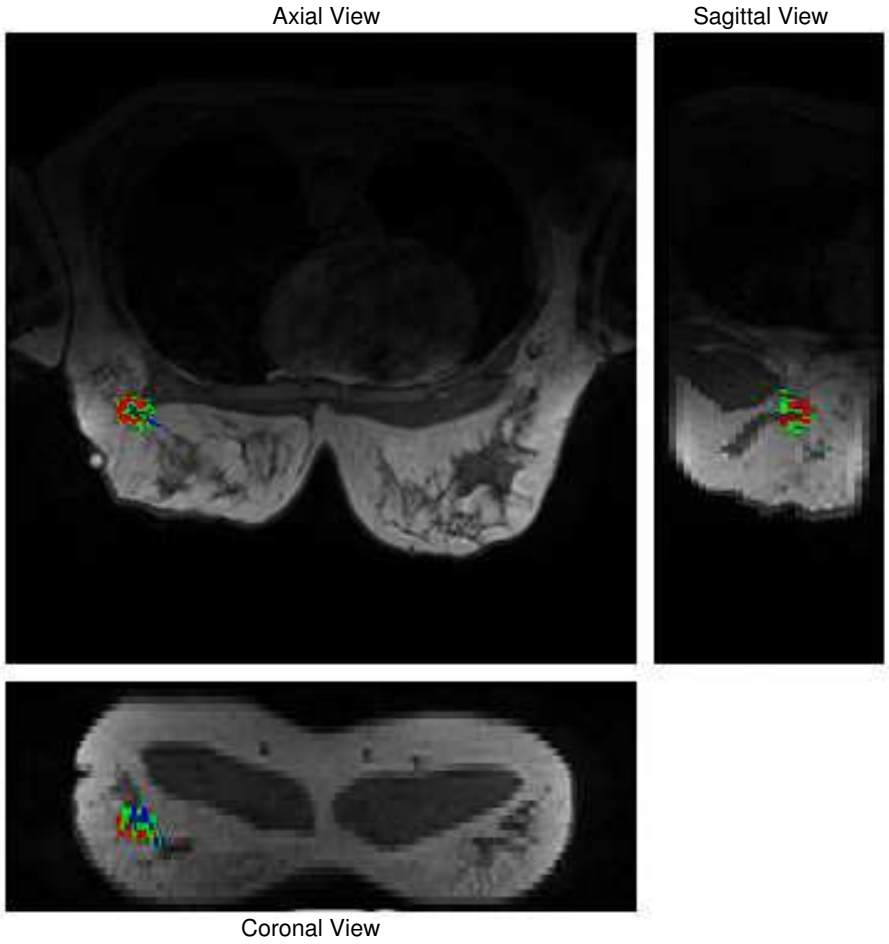


Figure 3.10: Image of 3TP evaluation of a recurrent ductal carcinoma. The 3TP pseudo-colours are superimposed on the precontrast image of the DCE-MRI sequence. The lesion was manually segmented by a radiologist, who marked voxels exhibiting significant intensity enhancements in the subtraction images. Subsequently, the marked lesion voxels were evaluated using 3TP. The erratic distribution of red, blue and green voxels illustrates the heterogeneity of cancerous tissue.

certain imaging protocol by evaluating the corresponding calibration maps. Thereby, the pseudo-colour at two positions B $= (0.3min^{-1}, 0.5)$ and M $= (0.95min^{-1}, 0.5)$ in the $(K^{\text{trans}}, v_{\text{e}})$-space are regarded particularly. The corresponding parameter tuples have shown to be typical for benign and malignant tissue, respectively. The pixel-mapping function is well calibrated, if the positions M and B are simultaneously located in the red and blue region, and both regions are separated by a green corridor of adequate width (Fig. 3.9, right).

Figure 3.10 shows axial, coronal and sagittal slices of a three-dimensional DCE-MR image. Voxels of the recurrent ductal carcinoma were manually labelled by a radiologist who marked significantly enhancing regions of tissue by means of subtraction images. Subsequently, the lesion voxels were coloured using the 3TP technique. Although the entire lesion is classified as a carcinoma according to the histopathologic examination, the distribution of lesion voxels

indicated as malignant (red), benign (blue) or suspicious (green) illustrates the heterogeneity of lesion tissue and a considerable number of voxels with indistinct wash-out characteristics. If lesions are examined with the 3TP technique, Kelcz et al., 2001 suggest to regard lesions exposing more than 15% red voxels as being malignant whereas lesions exposing more than 50% blue voxels and low intensities are likely to be benign.

3.5.3 Limitations of Model-Based Techniques

Model-based techniques provide an effective approach to analyse DCE-MRI signals with simple paradigms. Nevertheless, such paradigms require an explicit formulation of a reasonable mathematical model of the underlying physiological processes. Thus, their application is restricted on processes of low complexity, the investigator has reasonable knowledge of. Furthermore, each model makes a number of assumptions that may not be valid for every tissue or tumour type [Collins and Padhani, 2004].

Apart from requiring a mathematical model of the underlying process, fitting such a model to the measured data is a non-trivial task. Adaptation of the model parameters is computational expensive which deters the use of pharmacokinetic models at the workbench. The adapted model frequently does not fit the measured data exactly, leading to uncertainty about the reliability of the estimated parameter values. Additionally, indications for systematical overestimations of the K^{trans} parameters by commonly accepted models for DCE-MRI analysis have recently been reported [Collins and Padhani, 2004].

3.6 Data-Driven Image Analysis

Data-driven techniques based on *Artificial Neural Networks* (ANN) provide a basis for a more explorative approach to DCE-MRI data analysis. Unsupervised ANNs can directly be applied for the analysis of the DCE-MRI data recorded for the case which is currently under investigation. Thereby, a predefined number of clusters of signals is determined. Each cluster consists of temporal kinetic signals which are similar in the signal space according to a certain metric and may refer to a specific type of tissue such as benign or malignant. The outcome of the clustering process can be displayed as a three-dimensional image in which the colour of each voxel reflects the index of the prototype which is most similar to the corresponding temporal kinetic signal.

While unsupervised ANNs autonomously identify reasonable signal clusters from the data, supervised ANNs are applied in order to distinguish predefined signal classes, e.g. signals of normal and suspicious tissue. To this end, supervised ANNs correlate the signal information of the DCE-MR images with *label information* such as manual segmentations of lesions during an *adaptation* or *training* process. After adaptation of the ANN with the labelled data of a small cohort of cases, the trained ANN can be applied to infer the specified label, which is unknown for unseen cases, from the corresponding DCE-MRI sequences.

In contrast to model-based techniques which are limited to analysing temporal kinetic patterns, ANNs facilitate simultaneous processing of more general input patterns which may combine morphological features, features of the temporal dynamics of signals, multiparameter information derived from images with different T1/T2 weighting or texture information computed for small

image patches. On the other hand, the outcome of the data-driven techniques typically can not be interpreted in terms of physiologically meaningful quantities.

3.6.1 Applications of Supervised Artificial Neural Networks

Supervised ANNs, primarily *multilayer perceptrons* (MLP), have been employed for the evaluation of DCE-MRI data in different setups.

Lucht et al., 2001 employed a MLP to classify temporal kinetic signals of carcinoma, fibroadenoma, benign proliferative changes and parenchyma. Training data was collected by measuring average kinetic patterns of ROIs which were placed on compartments of lesions according to the parametric map obtained from a pharmacokinetic model. The trained classifier revealed 84% sensitivity and 81% specificity for the discrimination of benign and malignant signals but a poor performance for the subclassification of benign signals in fibroadenoma or benign proliferative changes. Additional experiments indicated an optimal performance for temporal sequences consisting of 28 measurements and a clearly reduced performance for sequences consisting of 3 measurements (78% sensitivity and 76% specificity). In [Lucht et al., 2002] the same setup was used for supervised segmentation of entire image volumes.

Tzacheva et al., 2003 utilised a MLP for classifying entire lesions as benign or malignant. The input patterns were features derived from a single static *contrast-enhanced magnetic resonance image* (CE-MRI) which was recorded with an imaging protocol with active fat suppression. Strong enhancing regions, i.e. potentially cancerous masses, were identified by a region-oriented segmentation of the contrast-enhanced image based on intensity thresholds. The segmented image was subsequently converted into a binary image. For each positive region, a pattern vector was computed combining mass margin and mass shape features in addition to simple texture features. These feature vectors were evaluated by a MLP to distinguish malignant regions from parenchyma and blood vessels. The system yielded 90% sensitivity and 91% specificity.

Abdolmaleki et al., 2001 trained a MLP to distinguish averaged kinetic signals of ROIs which were manually placed by a radiologist over malignant or benign tissue. The input pattern associated with each ROI consisted of quantitative features extracted from the kinetic signal such as the area-under-the-signal-curve, steepest slope in the wash-in part or the signal intensity after one, two and five minutes. Two additional features represented the age of the patient and the size of the associated ROI. A comparison of the ANN classification performance with those from a group of experienced radiologists and a group of low-experienced radiologists indicated that the ANN (97% sensitivity and 64% specificity) outperforms the latter and yields a performance comparable with an experienced radiologist.

The mentioned applications of supervised algorithms have in common that they either depend on dedicated imaging protocols [Tzacheva et al., 2003] or an extensive interaction with the user. The setup proposed by Lucht et al., 2002 requires the radiologist to manually place ROIs in the DCE-MR images in order to collect examples of tumour signals subsequently used for MLP training. A model-based technique (a pharmacokinetic model) is utilised for guiding the ROI placement, which in general is undesirable for the development of a second data-driven analysis setup. Abdolmaleki et al., 2001 propose a ANN based evaluation of average kinetic signals of ROIs, which were placed over the most enhancing region of each lesion. Since the same working step has to be performed for each new case, the application of this approach still depends on

interactions with the user. Additionally, the preceding working step, i.e. the localisation and delineation of the extent of the suspicious mass itself, is not considered in this work. In fact, results of a detailed comparison of learning algorithm based approaches for the localisation and delineation of suspicious masses with clinical standard procedures such as the manual evaluation of subtraction images have not yet been reported.

Apart from the shortcomings of their conceptional design, all most approaches have in common that they employ a multilayer perceptron. Even though the MLP is perhaps the most frequently used supervised classification algorithm in biomedical applications, recent advances in the area of machine learning and artificial neural networks have led to new supervised learning techniques which are easier to handle and are likely to achieve superior classification performance. For instance kernel-based techniques such as the *support vector machine* have shown impressive results in various applications. At the same time, they only require a small number of systematically tuneable hyperparameters as described in the following chapter.

3.6.2 Applications of Unsupervised Artificial Neural Networks

Several variants of unsupervised learning techniques have been employed for the purpose of visualisation and exploration of DCE-MRI data. These techniques solely utilise information as provided by the input patterns and reveal clusters of examples with similar signal characteristics or provide a compact display by transforming the high dimensional data.

Wismüller et al., 2002 and Meyer-Bäse et al., 2004 examined the application of *vector quantisation* (VQ) algorithms for learning of prototypes of temporal kinetic signals representing clusters in the signal space consisting of kinetic signals with similar characteristics. Wismüller et al., 2002 subsequently used the prototypes to segment the image volumes. Each voxel was labelled with the index of the prototype which is most similar to the associated temporal kinetic signal. The authors were able to segment the lesion mass from surrounding tissue as well as to find subdivisions of lesions in compartments with homogenous signal characteristic.

Jacobs et al., 2003 applied the *iterative self-organising data analysis* (ISODATA) algorithm, i.e. a *kmeans*-like VQ algorithm with a dynamic number of prototypes, for processing multiparameter DCE-MRI data. The input patterns were different combinations of static contrast-enhanced and non-enhanced MR images with a T1- or T2-weighting and optional fat-suppression. Based on the clustering result, a score was derived for the discrimination of malignant and benign lesions. Using the T1- and T2-weighted, enhanced and non-enhanced static images as features, the classification yielded 89% sensitivity and 74% specificity which is comparable to the performance of approaches evaluating the temporal kinetic signals of T1-weighted image sequences.

The determination of typical temporal kinetic signals which are more reliable than signals averaged over entire lesions was considered by Chen et al., 2004. A *fuzzy c-means* (FCM) clustering algorithm was applied for segmenting lesion masses in compartments of similar signals, each represented by an averaged temporal kinetic pattern with lower variance. The averaged temporal kinetic pattern exhibiting the strongest initial enhancement was used as a prototype for the lesion. For the following classification of the lesion as benign or malignant, the quantitative features maximum uptake, peak location, uptake rate and wash-out rate were calculated and evaluated by a *linear discriminant analysis* (LDA). The comparison of the classification performance with that yielded by a LDA evaluating the temporal kinetic pattern averaged over the entire lesion indicated

significantly increased performance of the FCM based classification scheme.

Yoo et al., 2002 applied *independent component analysis* (ICA) for the purpose of lesion detection and characterisation. The extracted independent components were interpreted as *reference waveforms* showing certain characteristics of the temporal kinetic signals. Subsequently, the DCE-MR sequence was displayed as a single three-dimensional grey value images with voxel intensities reflecting the correlation of the corresponding temporal kinetic pattern with a user-selected reference waveform. Since the ICA was able to extract reference waveforms with typical signal characteristics of malignant tissue, the malignant compartments of the lesions were highlighted in the corresponding correlation images.

3.7 Summary

DCE-MRI has proven to be a valuable imaging technique which is highly sensitive to the vascular changes of cancerous tissues. Beside morphological features, the key-information is the course of the concentration of contrast agent molecules in the tissue as measured in form of temporal sequences of three-dimensional MR images. Examination of the temporal kinetic signals enables the investigator to localise and characterise suspicious masses, but is a laborious and time-consuming task due to the multitemporal nature and the large amount of image data.

Computer aided diagnosis systems using model-based or data-driven techniques are currently under investigation. Model-based approaches are based on explicitly defined mathematical models of underlying physiological processes and demand for a-priori knowledge. Even though the pharmacokinetical models use several convenient simplifications, they provide valuable information with a clear histopathological meaning.

Data-driven approaches using supervised and unsupervised artificial neural networks have recently been identified as an appropriate alternative technique. ANN based systems derive implicit models of the temporal kinetic signals during data-driven adaptation processes. Thus, they do not rely on a-priori knowledge about the underlying processes and can be applied even if signal characteristics are complex.

In this thesis, modern machine learning techniques and ANNs are applied for the three main work tasks of DCE-MRI analysis, namely *efficient data visualisation*, *localisation of suspicious masses* and *characterisation of suspicious masses*. The goal is to develop applications based on ANNs which are *consistently data-driven* and do *not depend on preprocessing steps based on model-based techniques*. The proposed applications will only require data acquired during standard clinical diagnosis processes, which are geared towards the medical requirements of breast cancer diagnosis and not to the requirements of ANN applications.

4 Supervised Learning - Concepts, Algorithms and Evaluation

In this chapter, basic ideas of supervised learning algorithms are discussed. After a brief review of two principles of supervised learning, the supervised learning algorithms employed in this thesis are introduced. Techniques for evaluating the performance of classification models are described in the last section.

4.1 Concepts of Supervised Learning

Machine learning algorithms and *artificial neural networks* (ANN) are frequently applied to the task of classification, in which objects have to be assigned to a finite number of known categories or classes $\Omega = \{\omega_k\}, k = 1, \ldots, n_\Omega$. For classifying objects using an ANN, consistent vectorial descriptions of objects have to be provided. These vectorial descriptions may consist of three different types of features:

- *Quantitative features* reflecting continuous values e.g. the measurements of physical parameters.
- *Ordinal features* describing enumerated, discrete states of objects such as the grade of pathological tissue disorders.
- *Nominal features* describing the presence or absence of certain attributes e.g. 'smoker' or 'non-smoker' for clinical patients.

The collection of n_{in} features selected for the description of objects spans the *data space* $\mathcal{X} = \mathbb{R}^{n_{\text{in}}}$ and the vectorial description $\mathbf{x}_i$ of the i-th object can be regarded as a point in $\mathcal{X}$. Each point is attributed by a label $\mathbf{y}$ as a member of one of the considered classes ω_k. A simple, but nonetheless effective and frequently used method for classifying objects with unknown class labels is the comparison of the objects with a set Γ of examples with known class labels. The similarity of two objects $\mathbf{x}_i$ and $\mathbf{x}_j$ is measured by the distance $d(\mathbf{x}_i, \mathbf{x}_j)$ of two objects in $\mathcal{X}$ using a suitable metric. Thus, the *k-nearest-neighbour* (kNN) classifier [Hastie et al., 2001] designates an unlabelled object as a member of the majority class under its k closest examples $\mathbf{x}_j \in \Gamma$.

Contrary to the intuitive assumption that an increasing number of features facilitates discrimination of different object classes, artificial learning algorithms frequently suffer from high dimensional data spaces. Introduction of new features does not necessarily lead to a more effective discrimination of the objects. The information provided by new features may be completely irrelevant for the discrimination of the considered classes or is redundant, because the same information is already provided by other features. The phenomenon of degenerating performance of ANNs evaluating

high dimensional data is referred to in literature as *curse of dimensionality* [Bishop, 1995]. Since it is commonly not known in advance which subset of features provides discriminative information, two main concepts for dimensionality reduction exist.

Feature selection techniques either iteratively add features to an initially empty set of *active features* (*forward selection*) or iteratively remove a feature from the set of active features initially containing all features (*backward selection*). After each iteration, the *goodness* of the active set is evaluated by training the deployed classifier and evaluating its classification performance (*wrapper* method) or by evaluating a reasonable criterion independent of the classification task (*filter* method).

The second concept for avoiding the *curse of dimensionality* is to find a smaller subset of *meta features* $\bar{\mathbf{x}} \in \bar{\mathcal{X}}$, $\dim(\bar{\mathcal{X}}) < \dim(\mathcal{X})$ which are weighted linear or nonlinear combinations of the original features. Such features are obtained by optimisation of a reasonable criterion such as minimising the *least square* criterion

$$E = \sum_{i=1}^{N} \|\mathbf{x} - \mathbf{A}^T \bar{\mathbf{x}} + \boldsymbol{\mu}_{\mathbf{x}}\| \tag{4.1}$$

with the mean $\boldsymbol{\mu}_{\mathbf{x}}$ and mixing matrix $\mathbf{A}$ as in *principal component analysis* (PCA) [Jolliffe, 1986] or maximising the *mutual information* in *independent component analysis* (ICA) [Hyvärinen et al., 2001] for a set of N objects. These techniques belong to the group *unsupervised learning methods*, since they do not rely on labelled objects.

Once a suitable object description has been found, a supervised learning algorithm can be adapted using a set $\Gamma = \{(\mathbf{x}, y)_i\}, i = 1, \ldots, N$ of N labelled examples. Each example is a tuple consisting of the vectorial description $\mathbf{x}_i \in \mathcal{X}$ and the class label y_i. Depending on the number of classes to distinguish, the class membership can either be coded by a scalar $y_i = \{\pm 1\}$, as in the case of a binary data set, or by a n_Ω-dimensional vector $\mathbf{y}_i = \delta_{ij}$ using the *1-of-n_Ω* scheme with $\delta_{ij} = 1$ if $\mathbf{x}_i$ belongs to class ω_j and $\delta_{ij} = 0$ otherwise, as in the case of a multi-class data set.

4.1.1 Empirical Risk Minimisation

During adaptation of a supervised classification algorithm, the parameter set $\boldsymbol{\theta}$ of the *classification function*

$$f_{\boldsymbol{\theta}} : \mathcal{X} \mapsto \mathcal{Y}, \tag{4.2}$$

$$\mathbf{x} \to y(\mathbf{x}) \tag{4.3}$$

has to be learned from a set of labelled training data Γ sampled from the underlying data distribution $p(\mathbf{x}, y)$. A classification function $f_{\boldsymbol{\theta}}$ with parameters $\boldsymbol{\theta}$ adapted to a certain problem is also referred to as a *classification model*. The discrimination function associates each example $\mathbf{x} \in \mathcal{X}$ with a target vector $\mathbf{y}(\mathbf{x}) = f_{\boldsymbol{\theta}}(\mathbf{x})$ or target scalar $y(\mathbf{x}) = f_{\boldsymbol{\theta}}(\mathbf{x})$ indicating the assigned class $\omega \in \Omega$ according to the utilised coding scheme. One strategy for systematically adapting the parameters $\boldsymbol{\theta}$ is the *empirical risk minimisation* principle [Vapnik, 1995]. Thereby, the parameters

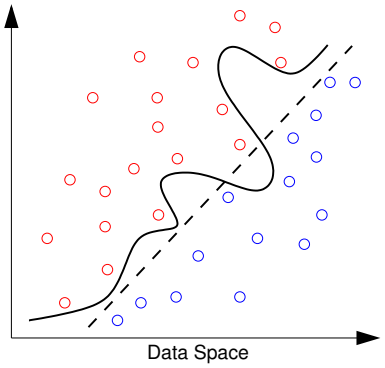


Figure 4.1: A classification model that minimises the empirical risk is not necessarily a minimiser of the expected risk. Both models (dashed and solid lines) depicted in the figure are able to separate the training examples of the two classes (red and blue circles) with no error. Nevertheless, it is unlikely that the model depicted by the solid line performs well on unseen examples, since the course of the decision boundary does not reflect the regularity of the data distribution reasonably well.

$\boldsymbol{\theta}$ are adapted to minimise the *empirical risk* or the *training error*

$$\mathcal{R}^{\Gamma}_{\text{emp}}[f_{\boldsymbol{\theta}}] = \frac{1}{N}\sum_{i=1}^{N} l(\mathbf{x}_i, y_i, f_{\boldsymbol{\theta}}) \tag{4.4}$$

measured on the training set Γ using a *loss-function* l. As an example, $\mathcal{R}^{\Gamma}_{\text{emp}}[f_{\boldsymbol{\theta}}]$ corresponds to the fraction of classification errors for a fixed set Γ of N examples with label $y \in \{\pm 1\}$ if the *zero-one loss function*

$$l(\mathbf{x}_i, y_i, f_{\boldsymbol{\theta}}) = \frac{1}{2}|y_i - f_{\boldsymbol{\theta}}(\mathbf{x}_i)| \tag{4.5}$$

is used. By minimising the empirical risk, the classifier learns a model of the underlying process generating the data. But a low empirical risk does not imply a low *expected risk*

$$\mathcal{R}_{\text{exp}}[f_{\boldsymbol{\theta}}] = \int l(\mathbf{x}_i, y_i, f) dP(\mathbf{x}, y) \tag{4.6}$$

for the underlying, typically unknown, joint probability distribution function $P(\mathbf{x}, y)$, although this *generalisation* on unseen data is the ultimate goal of learning. Nevertheless, the *law of large numbers* suggests that

$$\mathcal{R}^{\Gamma}_{\text{emp}}[f_{\boldsymbol{\theta}}] \underset{N\to\infty}{\to} \mathcal{R}_{\text{exp}}[f_{\boldsymbol{\theta}}], \tag{4.7}$$

which is the main motivation for the *empirical risk minimisation* (ERM) induction principle.

Bias-Variance Tradeoff

It turns out that straightforward minimisation of the empirical risk $\mathcal{R}^{\Gamma}_{\text{emp}}[f_{\boldsymbol{\theta}}]$ as suggested by the ERM principle is problematic. If the classification model $f_{\boldsymbol{\theta}}$ is allowed to be taken out of a

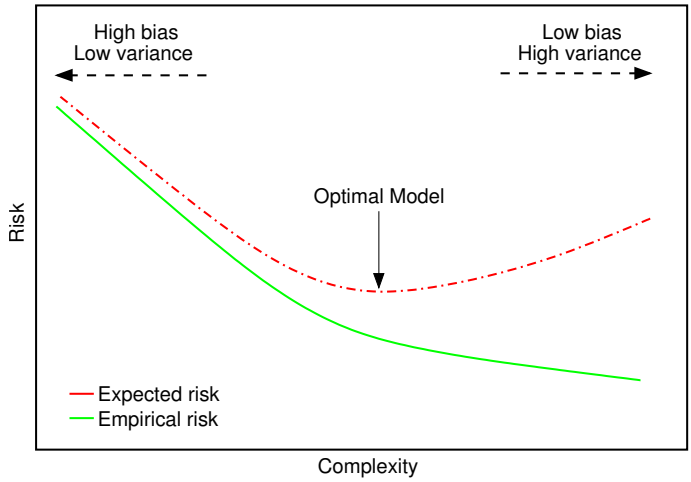


Figure 4.2: Bias-variance tradeoff for choosing a classifier of reasonable complexity. The complexity of a model should be selected such that it is sufficient to minimise the expected risk. Selecting a more complex model may lead to further reduction of the empirical risk but not of the expected risk. In this case, the model is said to *overfit* the training data.

family of functions of arbitrary complexity, there will be a large number of models that perfectly reproduce the class label for the given training set. However, this does not necessarily imply that these models carry information about the true data distribution and are able to infer the label for unseen examples equally well (Fig. 4.1).

Beside the empirical risk, the *complexity* or *capacity* of the model represented by a classifier has to be considered during training. A model of increasing complexity naturally leads to a decreasing $\mathcal{R}^{\Gamma}_{\text{emp}}[f_{\boldsymbol{\theta}}]$ but not necessarily to a decreasing $\mathcal{R}_{\text{exp}}[f_{\boldsymbol{\theta}}]$. If the model is too complex it may *overfit* the data, whereas a model which is not flexible enough may *underfit* the data (Fig. 4.2). This phenomenon is also known as the *bias-variance-tradeoff* [Hastie et al., 2001]. In consequence, the function $f_{\boldsymbol{\theta}}$ is usually constrained to e.g. a set of smooth decision boundaries since the underlying true decision boundary is often smooth in real-world problems.

Most classifiers are controlled by certain parameters which influence the complexity of the represented model like the number of nearest neighbours in the case of the kNN. These *hyperparameters*[1] are frequently selected manually by training a classifier with different parameterisations and approximating the corresponding prediction error on a separate test set of unseen examples by n-*fold cross-validation* or by the *leave-one-out error* [Hastie et al., 2001].

4.1.2 Structural Risk Minimisation

An alternative concept of learning is the *structural risk minimisation* (SRM) principle. The reader is referred to the book of Vapnik, 1995 for a detailed introduction. In SRM, learning aims at finding an appropriate trade-off between minimisation of empirical risk and model complexity. Model parameters are estimated by minimising an upper bound on the expected risk. With a

[1] The term *hyperparameters* refers all parameters of learning algorithms that have to be manually adjusted by the user.

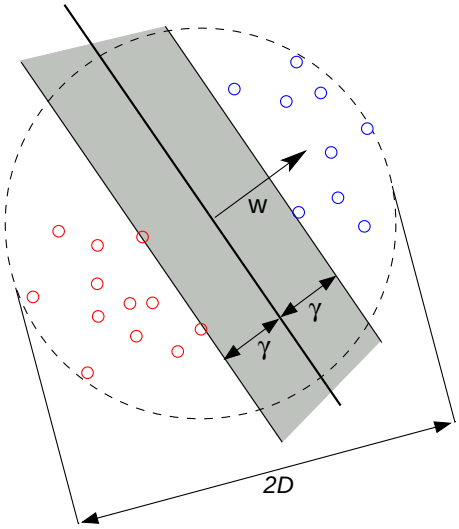


Figure 4.3: The VC dimension h of an optimal hyperplane can be bounded by the *radius-margin-bound*, i.e. the ratio of the radius D of the smallest sphere enclosing all training examples and the margin γ being the distance between the closest training examples and the hyperplane. Since D is constant for a given set of training examples, maximising the margin gives the smallest value for the bound on h.

probability of $1-\eta$ with $\eta \in [0;1]$, the classification model $f_{\boldsymbol{\theta}}$ that minimises the empirical risk $\mathcal{R}^{\Gamma}_{\text{emp}}[f_{\boldsymbol{\theta}}]$ on the given training set Γ satisfies

$$\mathcal{R}^{\Gamma}_{\text{exp}}[f_{\boldsymbol{\theta}}] \leq \mathcal{R}^{\Gamma}_{\text{emp}}[f_{\boldsymbol{\theta}}] + \sqrt{\frac{h}{N}\left(\log\frac{2N}{h}+1\right) - \frac{1}{N}\log\frac{\eta}{4}}, \tag{4.8}$$

where h is a quantity referred to as *Vapnik-Chernovenkis* (VC) dimension [Vapnik, 1995]. The VC dimension of a family of classifiers is a measure of the complexity of the space of all solutions that can be generated for any given data set. The right hand side of (4.8) is often referred to as the *VC bound*. A convenient property of the VC bound is that it is distribution-free, i.e. one does not need to know the distribution of the data to estimate the convergence rate of the learning algorithm, although tighter bounds could be derived if the data distribution functions were known.

For the family of hyperplanes, Vapnik showed that the VC dimension is bounded by the *radius-margin-bound*

$$h \leq \left(\frac{D}{\gamma}\right)^2, \tag{4.9}$$

with γ being the *margin* of the hyperplane, i.e. the distance of the hyperplane to the closest training example, and D being the radius of the smallest sphere enclosing all examples in Γ (Fig 4.3). This finding indicates that the *maximum margin hyperplane*, i.e. the hyperplane with the largest possible distance to the closest training example, gives the smallest value of the VC bound and can be expected to yield the highest generalisation power. The SRM principle is the theoretical foundation of the *support vector machine* algorithm [Vapnik, 1995], which combines the computation of maximum margin hyperplanes with implicit data transformations by nonlinear *Mercer's kernels* [Schölkopf et al., 1999a].

4.2 Support Vector Machine

In recent years, kernel based methods have gained much attention from the machine learning community. The most well-known kernel based algorithm is the (support vector machine) SVM [Vapnik, 1995, Schölkopf and Smola, 2002], which has shown superior performance for a wide range of binary classification tasks [Schölkopf and Smola, 2002, Bennett and Campbell, 2000, Cristianini and Shawe-Taylor, 2000]. The success of the SVM algorithm is based on several appealing properties. Firstly, training of the SVM follows the SRM principle and leads to a classification function which separates two classes ω_+ $(y = 1)$ and ω_- $(y = -1)$ by a *maximum margin hyperplane*. Since this hyperplane is the unique solution of a quadratic programme, the adaptation process is not sensitive to local minima. Secondly, introducing nonlinear kernel functions fulfilling the *Mercer's theorem* [Schölkopf and Smola, 2002, Schölkopf et al., 1999a] executes the algorithm implicitly in a new *feature space* $\mathcal{F}$ and leads to nonlinear discrimination functions in $\mathcal{X}$. Additionally, nonvectorial or structured data such as gene-sequences or text documents can be processed by the SVM as well as by other kernel-based methods using suitable kernel functions which embed the nonvectorial data in vectorial feature spaces [Shawe-Taylor and Cristianini, 2004, Meinicke et al., 2004, Leslie et al., 2002]. Finally, due to the availability of efficient adaptation algorithms and the low number of hyperparameters that have to be adjusted by the user, the SVM can be handled easily even by non-machine-learning experts.

4.2.1 Maximum Margin Hyperplanes

One of the most well-known algorithm for discriminating examples of two classes is the *perceptron* proposed by Rosenblatt, 1958. The perceptron is an iterative method for computing a hyperplane separating the examples of a binary labelled training set Γ. The adaptation process successively minimises the training error $\mathcal{R}^{\Gamma}_{\text{emp}}$. Novikoff, 1962 proved that a consistent solution is found after a finite number of iterations, given that the training examples of the two classes are linearly separable. Figure 4.4 illustrates the drawback of solely considering the training error during adaptation of the hyperplane. Both hyperplanes (bold black line and dashed gray line) in figure 4.4 are consistent solutions, i.e. they separate the two classes without misclassification of any training example. The final solution as found by the perceptron (dashed gray line) may vary significantly by reason of the large number of different but consistent hyperplanes that can be placed between the examples of the two classes. Nevertheless, the training set is just a limited representation of the underlying distribution $p(\mathbf{x}, y)$ and limited deviations $\Delta\mathbf{x}$ from the training samples have to be assumed for new examples. Therefore, considering the training error as well as the margin γ, i.e. the distance to the closest training example, during training leads to a *statistically more stable* discrimination function [Shawe-Taylor and Cristianini, 2004]. The classification outcome of the *maximum margin hyperplane* (bold black line), i.e. the hyperplane with the largest possible distance γ to the closest training example, does not change as long as $\|\Delta\mathbf{x}\| < \gamma$ holds.

Formally, the hyperplane $f_{\boldsymbol{\theta}}$ with $\boldsymbol{\theta} = (\mathbf{w}, b)$ is parameterised by a weight vector $\mathbf{w}$ and a bias b leading to the classification function

$$f_{\boldsymbol{\theta}}(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b.$$

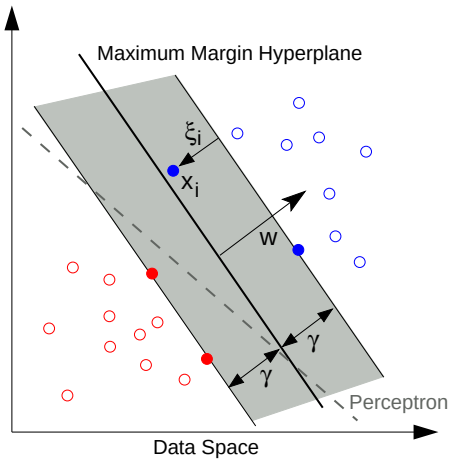


Figure 4.4: In contrast to a hyperplane found by the perceptron algorithm (dashed gray line), the maximum margin hyperplane (bold black line) maximises the distances to the closest training examples. This maximum margin hyperplane is solely defined by the support vectors (solid circles) located in the margin corridor depicted by the gray region. By introducing slack variables ξ_i, an adjustable amount of margin errors is tolerated in order to further increase the margin of the hyperplane.

The continuous outcome $f_{\boldsymbol{\theta}}(\mathbf{x})$, which is also referred to as *score* or *margin value* of $\mathbf{x}$, reflects the signed perpendicular distance of $\mathbf{x}$ to the hyperplane. The sign of the score indicates on which side of the hyperplane the example is located and thus corresponds to the assigned class label

$$y(\mathbf{x}) = \text{sgn}[f_{\boldsymbol{\theta}}(\mathbf{x})].$$

Without loss of generality, only canonical hyperplanes with

$$\min_{i=1,\ldots,N} |\langle \mathbf{w}, \mathbf{x}_i \rangle + b| = 1, \quad \mathbf{x}_i \in \Gamma$$

are considered in the following. Since the outcome for the closest training example is fixed to 1, the corresponding distance or the margin of the hyperplane is $\frac{1}{\|\mathbf{w}\|}$. Therefore, minimisation of the norm $\|\mathbf{w}\|$ maximises the margin γ of the hyperplane and adaptation of the SVM can be formulated as a constrained optimisation problem:

$$\min_{\mathbf{w},b} \quad \langle \mathbf{w}, \mathbf{w} \rangle \tag{4.10}$$

$$\text{subject to} \quad y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \quad \geq 1, \quad i = 1, \ldots, N. \tag{4.11}$$

Even though the solution $\boldsymbol{\theta}$ is statistically stable presuming γ is large, it is not robust in the sense that a single training example can significantly change the solution. Furthermore, a feasible solution only exists if the examples $\mathbf{x}_i \in \Gamma$ are linearly separable. In order to obtain a more robust solution, the constraints (4.11) are relaxed by introducing *slack variables* $\xi_i \geq 0$. The *soft margin*

formulation of the optimisation problem then becomes:

$$\begin{aligned} \min_{\mathbf{w},b} \quad & \langle \mathbf{w}, \mathbf{w} \rangle \quad + \quad C \sum_{i=1}^{N} \xi_i && (4.12) \\ \text{subject to} \quad & y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \quad \geq \quad 1 - \xi_i, \quad i = 1, \ldots, N && (4.13) \\ & \xi_i \quad \geq \quad 0, \quad i = 1, \ldots, N && (4.14) \\ & C \quad > \quad 0. && (4.15) \\ & && (4.16) \end{aligned}$$

The objective function (4.12) has now the form of a *regularised risk function* [Thikhonov and Arsenin, 1977]

$$\mathcal{R}^{\Gamma}_{\text{reg}}[f_{\boldsymbol{\theta}}] = \Omega[f_{\boldsymbol{\theta}}] + C\mathcal{R}^{\Gamma}_{\text{emp}}[f_{\boldsymbol{\theta}}],$$

with the *regularisation variable* C controlling the trade-off between the regularisation term $\Omega[f]$ controlling the complexity of the classifier and the empirical risk $\mathcal{R}^{\Gamma}_{\text{emp}}[f_{\boldsymbol{\theta}}]$ with the *soft-margin loss* function

$$l(\mathbf{x}_i, y_i, f_{\boldsymbol{\theta}}) = \max(0, 1 - y_i f_{\boldsymbol{\theta}}(\mathbf{x}_i)).$$

Depending on the value of C, an adjustable amount of margin errors ($y_i f_{\boldsymbol{\theta}}(\mathbf{x}_i) < 1$) or misclassified training examples ($y_i f_{\boldsymbol{\theta}}(\mathbf{x}_i) < 0$) is tolerated in order to increase the margin of the hyperplane.

The soft margin optimisation problem can be solved by adopting the *Lagrange theory* [Cristianini and Shawe-Taylor, 2000] leading to the corresponding *primal form*

$$L_P(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \mathbf{r}) = \frac{1}{2}\langle \mathbf{w}, \mathbf{w} \rangle + C \sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \alpha_i [y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) - 1 + \xi_i] - \sum_{i=1}^{N} r_i \xi_i.$$

Differentiating the primal form with respect to $\mathbf{w}, b$ and $\boldsymbol{\xi}$ imposing stationarity,

$$\begin{aligned} \frac{\partial L_P(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \mathbf{r})}{\partial \mathbf{w}} \quad &= \quad \mathbf{w} - \sum_{i=1}^{N} y_i \alpha_i \mathbf{x}_i = 0, && (4.17) \\ \frac{\partial L_P(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \mathbf{r})}{\partial \xi_i} \quad &= \quad C - \alpha_i - r_i = 0, && (4.18) \\ \frac{\partial L_P(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \mathbf{r})}{\partial b} \quad &= \quad \sum_{i=1}^{m} y_i \alpha_i = 0, && (4.19) \\ & && (4.20) \end{aligned}$$

and resubstituting the obtained relations into the primal form leads to the *dual form* of the objective function:

$$L_D(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \mathbf{r}) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} y_i y_j \alpha_i \alpha_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle.$$

The corresponding Karush-Kuhn-Tucker complementary conditions are

$$\begin{aligned} \alpha_i \Big[y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) - 1 + \xi_i \Big] &= 0, \quad i = 1, \ldots, m \\ \xi_i(\alpha_i - C) &= 0, \quad i = 1, \ldots, m. \end{aligned}$$

They imply that non-zero slack variables only occur for examples $\mathbf{x}_i$ whose distance is less than the desired margin $\frac{1}{\|\mathbf{w}\|}$, whereas $\mathbf{x}_i$ with $0 < \alpha_i < C$ lies at the target distance of $\frac{1}{\|\mathbf{w}\|}$.

In practice, the maximum margin hyperplane is computed by maximising the dual form of the objective function, because it is parameterised by a unique kind of parameters, the so-called *Lagrange coefficients* $\boldsymbol{\alpha}$. The optimal parameters $\boldsymbol{\alpha}$ are determined by solving the following constrained quadratic optimisation problem:

$$\max_{\boldsymbol{\alpha}} \quad \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} y_i y_j \alpha_i \alpha_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle \tag{4.21}$$

$$\text{subject to} \quad \sum_{i=1}^{N} y_i \alpha_i = 0 \tag{4.22}$$

$$0 \leq \alpha_i \leq, C, \quad i = 1, \ldots, N. \tag{4.23}$$

The classification function of the maximum margin hyperplane is then given by a weighted linear combination of examples $\mathbf{x}_i \in \Gamma$

$$f_{\boldsymbol{\theta}}(\mathbf{x}) = \sum_{i=1}^{m} y_i \alpha_i \langle \mathbf{x}_i, \mathbf{x} \rangle + b, \tag{4.24}$$

which is commonly defined by only a small number of training examples (the so called *support vectors*), since the majority of α_i tends to be zero. The class label of an unseen example $\mathbf{x}$ is still derived from the sign of the output $f_{\boldsymbol{\theta}}(\mathbf{x})$.

4.2.2 Kernel Functions

The introduction of the regularisation variable C controlling the trade-off between minimisation of model complexity and training error allows to compute an optimal hyperplane even for nonlinearly separable training data. To further extend the discriminative power of the linear classification method, input data are frequently mapped by a nonlinear transformation

$$\begin{aligned} \Phi : \mathcal{X} &\rightarrow \mathcal{F} \\ \mathbf{x} &\mapsto \Phi(\mathbf{x}) \end{aligned}$$

to a new, high-dimensional feature space $\mathcal{F}$, in which the linear algorithm will then be executed. A hyperplane in $\mathcal{F}$ corresponds to a nonlinear discrimination function in $\mathcal{X}$ by virtue of the nonlinear relationship between $\mathcal{X}$ and $\mathcal{F}$. Even if the training data are more likely to be linearly separable if they are transformed to a high dimensional feature space, two aspects have to be considered. The first aspect is the already mentioned *curse of dimensionality*, which can be

avoided by a reasonable regularisation of the maximum margin hyperplane. The second aspect is the increasing computational expense if the maximum margin hyperplane is explicitly computed in $\mathcal{F}$. The latter is alleviated by the application of the *kernel trick* which was first published by Aizerman et al., 1964 and allows for executing the SVM algorithm in $\mathcal{F}$ implicitly.

For an implicit execution of the SVM in a high or even infinite dimensional feature space which is nonlinear related to $\mathcal{X}$, a remarkable property of the corresponding quadratic programme given by (4.21)-(4.23) is exploited. Training examples $\mathbf{x}_i, \mathbf{x}_j \in \Gamma$ only appear as pairwise inner products, and the $N \times N$ matrix $\mathbf{I}$ with entries $I_{i,j} = \langle \mathbf{x}_i, \mathbf{x}_j \rangle$ provides sufficient information about the training data for adapting the SVM. In order to compute the maximum margin hyperplane in some $\mathcal{F}$ implicitly, the matrix $\mathbf{I}$ is replaced by a *kernel matrix* $\mathbf{K}$ with entries

$$K_{i,j} = K(\mathbf{x}_i, \mathbf{x}_j) = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle.$$

The entries $K_{i,j}$ are calculated by a computational inexpensive, nonlinear kernel function $K(\mathbf{x}_i, \mathbf{x}_j)$ and correspond to the inner product $\langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle$ in a feature space induced by some nonlinear transformation Φ, provided the kernel function fulfils *Mercer's condition* [Schölkopf et al., 1999a, Schölkopf and Smola, 2002]. After substituting the kernel function into (4.24), the discrimination function, i.e. the maximum margin hyperplane in $\mathcal{F}$, is given as a linear expansion of weighted kernel functions

$$f_{\boldsymbol{\theta}}(\mathbf{x}) = \sum_{i=1}^{N} y_i \alpha_i K(\mathbf{x}_i, \mathbf{x}) + b$$

and corresponds to a nonlinear discrimination function in $\mathcal{X}$. Frequently used kernel functions are the the *linear kernel*

$$K(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle, \tag{4.25}$$

the *polynomial kernel*

$$K(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle^p \tag{4.26}$$

which corresponds to a mapping Φ in the space of all monomials of degree p and the *Gaussian kernel*

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\Big(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\Big), \tag{4.27}$$

which allows for computing the maximum margin hyperplane in an infinite dimensional feature space [Cristianini and Shawe-Taylor, 2000]. Since the substitution of a kernel function $K(\mathbf{x}_i, \mathbf{x}_j)$ for the inner product $\langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle$ in the quadratic programme and in the discrimination function removes any appearance of Φ, the transformation Φ underlying the selected kernel function needs not be known explicitly.

4.2.3 Hyperparameter Selection

The discriminative power of the SVM depends crucially on the adjustment of hyperparameters such as the type and parametrisation of the kernel function and the regularisation parameter

C. Unless properties of the data suggest a specific kernel function, the function is commonly chosen by testing different standard kernels such as the linear kernel, the Gaussian kernel or the polynomial kernel on the training data. Although several approaches for automatic hyperparameter adjustment by optimising a suitable performance measure [Chapelle et al., 2002, Duan et al., 2003a, Keerthi, 2002] or by using genetic algorithms [Friedrichs and Igel, 2004] have been proposed, the standard approach for a low number of hyperparameters, e.g. for the two hyperparameters σ and C as in the case of SVMs with Gaussian kernel, is still a grid-search in the corresponding parameter space: Subsequently to the choice of a finite set of values for each hyperparameter the performance of the SVM for all combinations of hyperparameter values is evaluated either on a separate test set or by cross-validation on the training data. The combination of hyperparameter values yielding the best performance is then chosen as the optimal parametrisation for the given task.

4.2.4 Multi-Class Extensions

The basic version of the SVM algorithm considers only the case of binary classification, and there is no direct formulation of the SVM for the multi-class case in which examples of n_Ω classes have to be distinguished. Even though some formulations for solving multi-class problems by a single SVM-like optimisation problem have been proposed (see e.g. [Hsu and Lin, 2002]), decomposition schemes are still the most frequently used approaches for multi-class SVMs. Thereby, the multi-class problem is decomposed into a set of binary problems, each solved by a separate binary SVM. The outcomes of the individual binary classifiers are combined to a final multi-class prediction. This divide-and-conquer strategy of multi-class classification is not restricted to the SVM, but can also be used with every other binary classification algorithm. The most frequently used decomposition schemes are *one-versus-all* and *one-versus-one*, but other, more sophisticated approaches have been proposed (e.g. [Platt et al., 2000, Dietterich and Bakiri, 1995]).

One-Versus-All Decomposition

In the one-versus-all approach, n_Ω binary SVMs are trained. The k-th classifier is trained to discriminate examples of class ω_k from examples of the remaining classes. An unseen example $\mathbf{x}$ is then assigned to the class given by

$$\underset{k=1,\ldots,n_\Omega}{\operatorname{argmax}} f_{\boldsymbol{\theta}_k}(\mathbf{x}) \tag{4.28}$$

with $f_{\boldsymbol{\theta}_k}(\mathbf{x})$ as the signed margin value of the k-th binary SVM.

One-Versus-One Decomposition

In the one-versus-one scheme, a single binary classifier is trained for each possible pair of classes. Thus, the multi-class problem is decomposed into $\frac{n_\Omega(n_\Omega-1)}{2}$ binary problems. For the classification of an unseen example, the binary response $\operatorname{sgn}[f_{\boldsymbol{\theta}kl}(\mathbf{x})]$ of the classifier distinguishing examples from class ω_k versus examples from class ω_l is considered as a vote either for class ω_k or for ω_l. The example is finally assigned to the class which obtains the highest number of votes (*Max-Win* strategy).

Directed Acyclic Graph Support Vector Machine

An alternative approach for recombining the outputs of $\frac{n_\Omega(n_\Omega-1)}{2}$ binary classifiers is the *directed acyclic graph support vector machine* (DAGSVM) [Platt et al., 2000]. The DAGSVM is composed of a rooted binary directed acyclic tree with $\frac{n_\Omega(n_\Omega-1)}{2}$ internal nodes. Each node corresponds to one of the classification models discriminating examples of two classes. For the classification of an unseen example, the graph is traversed starting from the root node. At each node, the next subgraph is selected according to the output of the classification model corresponding to the current node. Thus, for the calculation of the finally assigned class label as given by the final leaf, only a subset of the classifiers has to be applied to the unseen example, which reduces the evaluation time.

Error-Correcting-Output-Codes Framework

The one-versus-all and one-versus-one schemes can be regarded as a special case of the more general *error-correcting-output-codes* (ECOC) framework proposed by Dietterich and Bakiri, 1995. In its early version, the multi-class problem is decomposed by adapting n_K classifiers on different partitions of the training data. Subsequently, the binary responses of the n_K binary classification models for an unseen example $\mathbf{x}$ are combined to a n_K-dimensional output vector $\mathbf{o}$. This vector is evaluated by a *decoding matrix* $\mathbf{D} \in \{\pm 1\}^{n_\Omega \times n_\mathsf{K}}$ containing an unique code vector for each class. The final class response is determined by calculating the best matching code vector using

$$\underset{k=1,\ldots,n_\Omega}{\operatorname{argmin}} \; d(\mathbf{D}_{k\bullet}, \mathbf{o}) \tag{4.29}$$

with $d(\mathbf{D}_{k\bullet}, \mathbf{o})$ as the *Hamming* distance between $\mathbf{o}$ and the k-th row of $\mathbf{D}$ containing the code vector of class ω_k. Later, the ECOC scheme was extended by [Allwein et al., 2000] to take the continuous margin values into account. For this purpose, the Hamming-based decoding was replaced by a decoding which employs a suitable loss-function.

In [Hsu and Lin, 2002] the different decomposition schemes for multi-class classification using binary SVMs are compared by means of standard benchmark data sets. Even though a significantly larger number of binary SVMs has to be trained for the one-versus-one and DAGSVM scheme, the training time as well as the evaluation time of unseen examples can be shorter than for the one-versus-all scheme. The computational complexity of the SVM training scales about quadratic to cubic with the number of training examples [Schölkopf et al., 1999b]. Thus, solving a larger number of smaller quadratic programmes can be computational less expensive than solving a smaller number of larger quadratic programmes. The computational expense of classifying an unseen example is dominated by the number of kernel evaluations, which depends on the number of support vectors. In the case of nonlinear kernel functions like the Gaussian kernel the number of support vectors and, therewith, the number of kernel evaluations needed for classifying an example typically increases with the number of training examples. For the multi-class data sets considered by Hsu and Lin, 2002, the total number of support vectors of the one-versus-one solution was smaller than for the one-versus-all solution resulting in an increased evaluation time for the latter scheme. In terms of accuracy, the experiments indicated comparable performance of one-versus-one and one-versus-all schemes with Gaussian kernel, but a slightly superior performance of the former scheme if the problem is solved by a set of SVMs with linear kernel.

4.2.5 Output Calibration

Class membership probability estimates, i.e. a-posteriori probabilities $P(\omega_k|\mathbf{x})$, are frequently desired as output from a classifier in order to combine the classification outcome with additional information sources for decision making. The SVM as well as many other classifiers only returns margin values or scores $f_{\boldsymbol{\theta}}(\mathbf{x}_i), f_{\boldsymbol{\theta}}(\mathbf{x}_j) \in [-\infty;\infty]$ describing the confidence that the examples $\mathbf{x}_i, \mathbf{x}_j$ belong to class ω_+. These score values allow for ranking of the examples from the least probable to the most probable members of class ω_+ under the assumption that $P(\omega_+|\mathbf{x}_i) < P(\omega_+|\mathbf{x}_j)$ whenever $f_{\boldsymbol{\theta}}(\mathbf{x}_i) < f_{\boldsymbol{\theta}}(\mathbf{x}_j)$. In order to obtain accurate estimates of a-posteriori probabilities from the SVM, the scores have to be calibrated by a suitable transformation function.

Calibrated Outputs for Binary SVMs

The simplest approach for obtaining a-posteriori probabilities from the binary SVM algorithm is by binning the range of scores as it is recommended for naive Bayes classifiers by Zadrozny and Elkan, 2001. Thereby, the a-posteriori probability $P(\omega_+|\mathbf{x})$ of an unseen example $\mathbf{x}$ is estimated by the fraction of positive training examples within the bin containing the corresponding score $f_{\boldsymbol{\theta}}(\mathbf{x})$.

Platt, 2000 proposed a post-processing of the SVM outcome $f_{\boldsymbol{\theta}}(\mathbf{x})$ by a scaling function. Driven by the fact that the relation between the empirical class membership probability $P(\omega_+|f(\mathbf{x}) = s)$, i.e. the number of training examples $\mathbf{x}$ with assigned score $f(\mathbf{x}) = s$ divided by the total number of training examples, and the score appears to follow the course of a sigmoid function for many data sets, the parametric model

$$P(\omega_+|\mathbf{x}) = \frac{1}{1 + \exp[w_1 f_{\boldsymbol{\theta}}(\mathbf{x}) + w_0]} \tag{4.30}$$

of a sigmoid function is utilised for postprocessing. The model parameters w_0 and w_1 are computed by minimising the *negative log-likelihood* [Bishop, 1995] using the *Levenberg-Marquardt* algorithm [Press et al., 1986]. Recently, the approach was numerically improved by Lin et al., 2003 who also replaced the Levenberg-Marquardt method by a simple Newton's method with backtracking line search for parameter optimisation.

Calibrated Outputs for Multi-Class SVMs

Several approaches have been proposed in order to obtain a-posteriori probabilities from a multi-class SVM. In [Duan et al., 2003b] a softmax function

$$P(\omega_k|\mathbf{x}) = \frac{\exp(w_k f_{\boldsymbol{\theta}_k}(\mathbf{x}) + w_{k0})}{\sum_{k=1}^{n_\Omega} \exp(w_k f_{\boldsymbol{\theta}_k}(\mathbf{x}) + w_{k0})} \tag{4.31}$$

is used for combining the margin values $f_{\boldsymbol{\theta}_k}(\mathbf{x})$ of the n_Ω binary SVMs of an one-versus-all multi-class SVM. The parameters of the softmax layer are selected by optimising a regularised form of the negative log-likelihood leading to the constrained optimisation problem:

$$\begin{aligned} \min \quad & E_{\mathsf{NLL}} = \tfrac{1}{2}\|\mathbf{w}\|^2 - \lambda \textstyle\sum_{i=1}^{m} \log[P(\omega = \mathbf{y}_i|\mathbf{x}_i)] && (4.32) \\ \text{subject to} \quad & w_k, w_{k0} > 0, k = 1, \ldots, n_\Omega, && (4.33) \end{aligned}$$

with $\|\mathbf{w}\|^2 = \sum_{k=1}^{n_\Omega}(w_k^2 + w_{k0}^2)$. For the one-versus-one multi-class SVM, the softmax layer is adjusted according to

$$P(\omega_k|\mathbf{x}) = \frac{\exp(\sum_{l\neq k} w_{kl} f_{\boldsymbol{\theta}_{kl}}(\mathbf{x}) + w_{k0})}{\sum_{k=1}^{n_\Omega} \exp(\sum_{l\neq k} w_{kl} f_{\boldsymbol{\theta}_{kl}}(\mathbf{x}) + w_{k0})} \tag{4.34}$$

with $f_{\boldsymbol{\theta}_{kl}}(\mathbf{x})$ as the margin value discriminating examples of class ω_k versus ω_l. In [Wu et al., 2004], a-posteriori probabilities are obtained from a one-versus-one multi-class SVM by *pairwise-coupling* [Hastie and Tibshirani, 1998].

Parameter Fitting

For fitting the parameters of the scaling function, the data set $\{(f_{\boldsymbol{\theta}}(\mathbf{x}_i), \mathbf{y}_i)\}$ has to be carefully selected in order to avoid over-fitting. The margin values $f_{\boldsymbol{\theta}}(\mathbf{x}_i), \mathbf{x}_i \in \Gamma$ of the training set Γ used for adaptation of the SVM only provide substantially biased estimates of the distribution of score value for examples beyond the margin (non support vectors). The training examples near the margin with $(\mathbf{y}_i f_{\boldsymbol{\theta}}(\mathbf{x}_i) \approx 1)$ are forced to take an absolute value of 1. Examples that fail the margin with $(1 - \mathbf{y}_i f_{\boldsymbol{\theta}}(\mathbf{x}_i) > 0)$ are also biased as they are pushed towards the margin by the corresponding Lagrange coefficients α_i. Thus, Platt, 2000 suggests to fit the parametric model either to the score values calculated for examples of a hold out $\bar{\Gamma} \subset \Gamma$ disjoint to the training set used for adapting the SVM or to the score values obtained from a n-fold-cross-validation process on the training data. To this end, the training set is divided into n disjoint subsets. After adapting the SVM with the union of $n-1$ subsets, the score values are calculated for examples of the unseen subset. By repeating this process for all permutations of the n subsets, significantly less biased score values can be obtained for the entire training set.

4.3 Linear Discriminant Analysis

Linear discriminant analysis (LDA) is a classical statistical approach for discriminating examples of two or more classes by hyperplanar decision boundaries for which computation different but related techniques have been proposed in the past (see [Hastie and Tibshirani, 1998]). One of the most well-known method is the *Fisher's linear discriminant* [Bishop, 1995] which regards the task of discriminating examples of the two classes ω_+, ω_- as a dimensionality reduction problem. Examples $\mathbf{x}_i \in \Gamma$ are mapped to a one-dimensional space by projecting them onto a vector $\mathbf{w}$ (Fig 4.5). The sign of the classification function

$$f_{\boldsymbol{\theta}}(\mathbf{x}_i) = w_0 + \langle \mathbf{x}_i, \mathbf{w} \rangle$$

with $\boldsymbol{\theta} = (\mathbf{w}, w_0)$ provides a class label according to $y(\mathbf{x}_i) = \text{sgn}[f_{\boldsymbol{\theta}}(\mathbf{x}_i)]$. Therewith, the classification function corresponds to a hyperplane in $\mathcal{X}$ with a perpendicular distance of w_0 to the origin. The vector $\mathbf{w}$ determined by the Fisher's linear discriminant maximises the *Rayleigh coefficient*

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{B} \mathbf{w}}{\mathbf{w}^T \mathbf{W} \mathbf{w}},$$

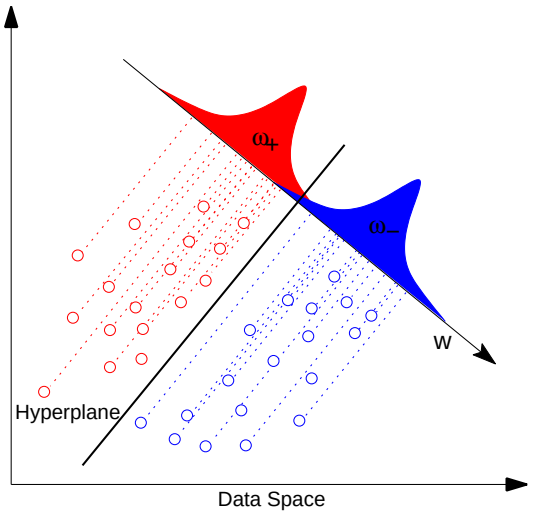


Figure 4.5: The normal vector $\mathbf{w}$ perpendicular to the hyperplane of the Fisher's linear discriminant lies in the direction that maximises the separability of the examples of class ω_+ (red) and ω_- (blue) after projection onto $\mathbf{w}$.

measuring the ratio of the *between-class scatter* $\mathbf{w}^T\mathbf{B}\mathbf{w}$ and the *within-class scatter* $\mathbf{w}^T\mathbf{W}\mathbf{w}$ for a training set Γ. The *between-class scatter matrix* $\mathbf{B}$ is given by

$$\mathbf{B} = (\boldsymbol{\mu}_- - \boldsymbol{\mu}_+)(\boldsymbol{\mu}_- - \boldsymbol{\mu}_+)^T$$

with the class-specific sample mean $\boldsymbol{\mu}_+, \boldsymbol{\mu}_-$ computed by

$$\boldsymbol{\mu}_i = \frac{1}{N_{\omega_i}} \sum_{\mathbf{x} \in \omega_i} \mathbf{x}_i.$$

The *within-class scatter matrix* $\mathbf{W}$ is defined as

$$\mathbf{W} = \mathbf{W}_+ + \mathbf{W}_-$$

with the class-specific scatter matrices

$$\mathbf{W}_k = \sum_{\mathbf{x}_i \in \omega_k} (\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^T.$$

The vector $\mathbf{w}$ maximising $J(\mathbf{w})$ can then be derived from the eigenvalue problem

$$\mathbf{W}^{-1}\mathbf{B}\mathbf{w} = \lambda\mathbf{w}$$

with some constant λ. Due to the fact that in this particular case $\mathbf{B}\mathbf{w}$ lies always in the direction of $(\boldsymbol{\mu}_- - \boldsymbol{\mu}_+)$ and the scale factor is irrelevant, the optimal $\mathbf{w}$ can also be calculated by

$$\mathbf{w} = \mathbf{W}^{-1}(\boldsymbol{\mu}_- - \boldsymbol{\mu}_+)$$

leading to the classification function

$$f_{\boldsymbol{\theta}}(\mathbf{x}) = w_0 + \mathbf{x}^T \mathbf{W}^{-1}(\boldsymbol{\mu}_- - \boldsymbol{\mu}_+). \tag{4.35}$$

The constant shift w_0 can be obtained by minimising the empirical error on the training set Γ. Alternatively, w_0 can be calculated directly, whenever the class-conditional densities are multivariate normal

$$p(\mathbf{x}|\omega_k) = \frac{1}{(2\pi)^{\frac{n_{\text{in}}}{2}} \det(\Sigma_k)^{\frac{1}{2}}} \exp\Big(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_k)^T \Sigma_k^{-1}(\mathbf{x} - \boldsymbol{\mu}_k)\Big),$$

with equal covariance matrices $\Sigma = \Sigma_+ = \Sigma_-$ for both classes ω_+ and ω_- [Hastie et al., 2001]. In case of a two class problem, the class label can be determined by means of the sign of the log-ratio

$$\begin{aligned} \log\frac{P(\omega_-|\mathbf{x}_i)}{P(\omega_+|\mathbf{x}_i)} &= \log\frac{P(\omega_-)}{P(\omega_+)} + \log\frac{p(\mathbf{x}_i|\omega_-)}{p(\mathbf{x}_i|\omega_+)} \\ &= \log\frac{P(\omega_-)}{P(\omega_+)} - \frac{1}{2}(\boldsymbol{\mu}_- + \boldsymbol{\mu}_+)^T \Sigma_k^{-1}(\boldsymbol{\mu}_- - \boldsymbol{\mu}_+) \\ &\quad + \mathbf{x}_i^T \Sigma^{-1}(\boldsymbol{\mu}_- - \boldsymbol{\mu}_+). \end{aligned} \tag{4.36}$$

Since $\mathbf{W}$ is proportional to the sample covariance matrix Σ of the entire training set Γ [Bishop, 1995], (4.36) is equivalent to (4.35) with

$$w_0 = \log\frac{P(\omega_-)}{P(\omega_+)} - \frac{1}{2}(\boldsymbol{\mu}_- + \boldsymbol{\mu}_+)^T \Sigma_k^{-1}(\boldsymbol{\mu}_- - \boldsymbol{\mu}_+). \tag{4.37}$$

This finding suggests choosing w_0 such that the hyperplane is shifted to the position where the a-posteriori probabilities of both classes are equal.

A related discrimination function can be formulated by using *Bayes' theorem* and explicit modelling of the class-conditional densities. The a-posteriori probability of membership of class ω_+ is given by

$$P(\omega_+|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_+)P(\omega_+)}{p(\mathbf{x}|\omega_+)P(\omega_+) + p(\mathbf{x}|\omega_-)P(\omega_-)}.$$

If the class-conditional densities $p(\mathbf{x}|\omega_+)$ and $p(\mathbf{x}|\omega_-)$ are modelled as normal distributions with equal covariance matrix and examples $\mathbf{x}$ are assigned to the class having the largest a-posteriori probability then both classes are separated by a hyperplanar decision boundary. In contrast to the Fisher's linear discriminant, the output is within a fixed range of $[0; 1]$ and can be interpreted in terms of probabilities.

4.4 Local Sigmoid Map

The *local sigmoid map* (LSM) is a supervised hierarchical artificial neural network which employs a *divide-and-conquer* scheme [Jacobs et al., 1991] for solving binary and multiclass classification tasks. The discrimination function

$$\begin{aligned} f_{\boldsymbol{\theta}} : \mathcal{X} &\mapsto \mathcal{Y} = [0;1]^{n_\Omega} \\ \mathbf{x} &\to \mathbf{y}(\mathbf{x}) \end{aligned}$$

associates each example $\mathbf{x} \in \mathcal{X}$ with a target vector $\mathbf{y}(\mathbf{x}) = f_{\boldsymbol{\theta}}(\mathbf{x})$ indicating the assigned class by the largest component according to the 1-of-n_Ω scheme. The function is modelled by subdividing a complex classification task into a number of less complex partial tasks, each requiring a classification function of lower complexity. The entire training process consists of two steps:

1. Approximation of the distribution of the pooled training examples by a set of prototypes $\{\mathbf{w}_l\}, \mathbf{w}_l \in \mathcal{X}, \quad l = 1, \ldots, n_\mathcal{V}$ as computed by an arbitrary *Vector Quantisation* (VQ) algorithm [Hastie et al., 2001]. The resulting prototypes describe a partition of the data space in $n_\mathcal{V}$ subregions or *voronoi cells* according to

$$\mathcal{V}_l = \{\mathbf{x} \in \mathcal{X} : \|\mathbf{w}_l - \mathbf{x}\| < \|\mathbf{w}_m - \mathbf{x}\|, \forall l \neq m\}, \quad l, m = 1, \ldots, n_\mathcal{V}.$$

2. Supervised adaptation of a local expert for each voronoi cell. Thereby, the local expert associated with the l-th voronoi cell $\mathcal{V}_l$ is adapted for the classification function

$$\begin{aligned} \tilde{f}_l : \mathcal{V}_l &\mapsto \mathcal{Y} \\ \mathbf{x} &\to \mathbf{y}(\mathbf{x}), \end{aligned}$$

   which maps each example $\mathbf{x} \in \mathcal{V}_l$ to the target vector $\mathbf{y}(\mathbf{x})$ by a single-layer neural network with softmax activation function:

$$y_k(\mathbf{x}) = \frac{\exp(\boldsymbol{\beta}_k^T \mathbf{x})}{\sum_{k=1}^{n_\Omega} \exp(\boldsymbol{\beta}_k^T \mathbf{x})}, \quad k = 1, \ldots, n_\Omega.$$

Thus, the LSM is formally defined by the set

$$\Theta = \left\{ \left(\mathbf{w}, \{\boldsymbol{\beta}_k\}\right)_l \right\}, \quad k = 1, \ldots, n_\Omega, \quad l = 1, \ldots, n_\mathcal{V}$$

of $n_\mathcal{V}$ tuples $(\mathbf{w}, \{\boldsymbol{\beta}_k\})$, each defining a single local expert. The prototype $\mathbf{w} \in \mathcal{X}$ defines the location and subregion associated with the expert. The set $\{\boldsymbol{\beta}_k\}$ is the parametrisation of the corresponding layer of n_Ω neurons with softmax activation function. For notational simplicity it is assumed that a constant component x_0 is appended to each example $\mathbf{x}$ before being exposed to the neuron layer. Thereby, the bias term β_0 is incorporated into each parameter vector $\boldsymbol{\beta} \in \mathbb{R}^{n_{\text{in}}+1}$.

This architecture of an ANN has been realised in different variants before. In [Ritter et al., 1992], the *local linear map* (LLM) algorithm is proposed which combines *online-kmeans vector quantisation* with linear regression functions. Martinetz et al., 1993 replaced the *online-kmeans* vector quantisation algorithm by the more robust *neural gas* algorithm. While the application of linear regression functions allows to approximate functions $f : \mathcal{X} \mapsto \mathcal{Y} = \mathbb{R}^{n_{out}}$, the LSM uses single-layer neural networks with softmax activation functions for the classification of examples. Each example is mapped to a n_Ω-dimensional target vector $\mathbf{y}(\mathbf{x}) \in \mathcal{Y} = [0;1]^{n_\Omega}$ with the k-th

component $y_k(\mathbf{x})$ indicating the a-posteriori probability $P(\omega_k|\mathbf{x})$, and the according class is given by

$$\underset{\omega_k \in \Omega}{\text{argmax}} P(\omega_k|\mathbf{x}).$$

This modification of the LLM architecture simplifies further processing of the outcome, since each component of $\mathbf{y}(\mathbf{x})$ has a well-defined range. For instance in chapter 6, the outcome of the LSM classifying malignant, normal and benign tissue can directly be interpreted as a RGB colour and utilised for computing pseudo-colour images of DCE-MRI data.

4.4.1 Prototype Adaptation

For the approximation of the distribution of training examples $(\mathbf{x}, \cdot) \in \Gamma$ in $\mathcal{X}$ in terms of vector quantisation, the prototypes $\{(\mathbf{w}, \cdot)_l\}$ are adapted using the *neural gas* (NG) algorithm [Martinetz et al., 1993]. The NG algorithm proved to be efficient and robust against different initialisation conditions. In contrast to the *kmeans* algorithm, which only adapts the prototype which is closest to the presented example according to a certain metric, the NG algorithm adapts a certain number of prototypes considering their rank order to overcome local minima.

After initialisation of the prototypes with a random subset of the training data, the distortion error is minimised by a stochastic gradient descent. For each example $(\mathbf{x}, \cdot) \in \Gamma$ randomly selected at iteration step t, the rank $r_l(\mathbf{x}) \in \{0, 1, \ldots, n_\Theta - 1\}$ of each prototype $\mathbf{w}_l$ is calculated according to the distance $d(\mathbf{x}, \mathbf{w}_l) = \|\mathbf{x} - \mathbf{w}_l\|$ with $r = 0$ as the rank of the closest prototype. Afterwards, each prototype $\mathbf{w}_l$ is adapted according to

$$\Delta \mathbf{w}_l = \epsilon_t \cdot h(\sigma_t, r_l) \cdot (\mathbf{x} - \mathbf{w}_l), \tag{4.38}$$

with learning rate ϵ_t and the neighbourhood function

$$h(\sigma_t, r_l) = \exp\Big(-\frac{r_l}{\sigma_t}\Big), \tag{4.39}$$

which decreases with each iteration step t. After adapting the prototypes, $\mathcal{X}$ is tessellated in $n_\mathcal{V}$ voronoi cells using the *winner-takes-all* (WTA) rule

$$\mathbf{w}^*(\mathbf{x}) = \underset{(\mathbf{w}, \cdot) \in \Theta}{\text{argmin}} \|\mathbf{x} - \mathbf{w}\|, \tag{4.40}$$

which assigns each example $\mathbf{x} \in \mathcal{X}$ to its closest prototype $\mathbf{w}^*(\mathbf{x})$.

4.4.2 Local Expert Adaptation

After approximating the distribution of the pooled training data, the parameters $\{(\cdot, \{\boldsymbol{\beta}_k\}_l\}$ of the $n_\mathcal{V}$ single-layer ANNs need to be trained in order to map examples $\mathbf{x}$ to the output value $\mathbf{y}(\mathbf{x})$. To this end, the empirical risk as measured by the *cross-entropy* error [Bishop, 1995]

$$l(\mathbf{x}_i, \mathbf{y}_i, f) = -\sum_{k=1}^{n_\Omega} y_{ik} \ln[y_k(\mathbf{x}_i)] \tag{4.41}$$

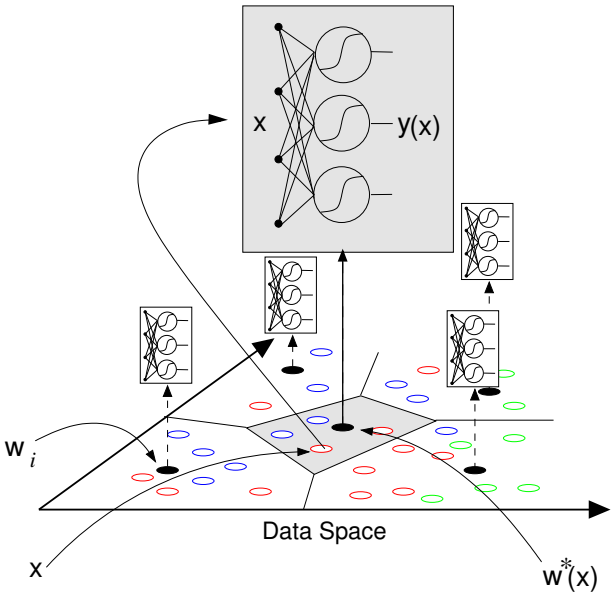


Figure 4.6: Evaluation of an example $\mathbf{x}$ by a LSM operated in the hard-gating mode. First, the best matching prototype $\mathbf{w}^*(\mathbf{x})$ is determined. The output $\mathbf{y}(\mathbf{x})$ is then calculated by presenting the examples to the local expert associated with $\mathbf{w}^*(\mathbf{x})$.

is minimised using stochastic gradient descent. For each example $(\mathbf{x}, \mathbf{y})_i \in \Gamma$ randomly selected at iteration step t, the closest prototype $\mathbf{w}^*(\mathbf{x}_i)$ is determined using the WTA-rule (4.40). The parameter vectors $\{\boldsymbol{\beta}_k\}$ associated with $\mathbf{w}^*(\mathbf{x}_i)$ are adapted according to

$$\Delta\boldsymbol{\beta}_k = -\eta_t \cdot \Big(y_k(\mathbf{x}_i) - y_{ik}\Big) \cdot \mathbf{x}_i, \quad k = 1, \ldots n_\Omega \tag{4.42}$$

with a learning rate η_t which decreases with each iteration step t.

4.4.3 Classification of Unseen Examples

Figure 4.7 depicts a solution of the LSM ($n_\mathcal{V} = 6$) for a binary data set in which each class is generated by two Gaussian distributions. The left image shows the distribution of training examples $(\mathbf{x}, \mathbf{y})_i \in \Gamma$ for the classes ω_1 (red crosses) and ω_2 (blue circles). The locations of the six prototypes are marked by solid black diamonds. The $y_1(\mathbf{x}) = y_2(\mathbf{x})$ or $P(\omega_1|\mathbf{x}) = P(\omega_2|\mathbf{x})$ isoline is indicated as a black line. The right image exposes the same isoline in addition to the colour coded a-posteriori probability $P(\omega_1|\mathbf{x})$. For the shown solution, the LSM is operated in the *hard gating mode*: The VQ layer acts as a hard gating network and directs the input pattern to a single local expert using the WTA-rule. The outcome $\mathbf{y}(\mathbf{x})$ is then solely determined by the ANN associated with the best matching prototype $\mathbf{w}^*(\mathbf{x})$ (see Fig. 4.6). Since each neuron of the single-layer ANN distinguishes the examples by a linear function, the collectivity of local experts leads to a piecewise-linear classification function $f_{\boldsymbol{\theta}}(\mathbf{x})$ (black line). In the hard-gating operation mode, improper strong variations of the outcome may appear at the contact points of two voronoi cells indicated by the discontinuities in the colours of certain subregions of $\mathcal{X}$. Even though two

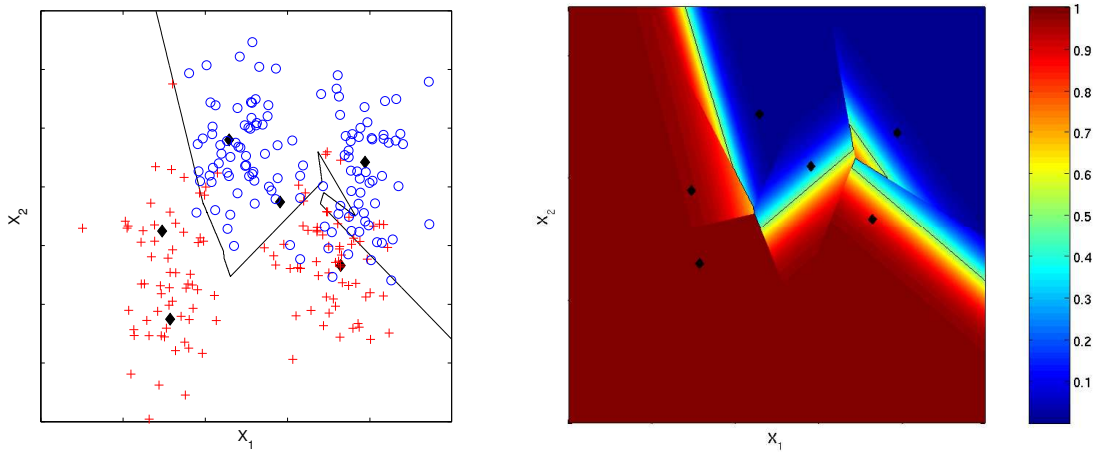


Figure 4.7: Results of a LSM operated in the hard-gating mode, for the synthetic two class data set used in [Ripley, 1996]. Left: Distribution of training examples of the classes ω_1 (red crosses) and ω_2 (blue crosses). The black line indicates the $y_1(\mathbf{x}) = y_2(\mathbf{x})$ or $P(\omega_1|\mathbf{x}) = P(\omega_2|\mathbf{x})$ isoline. Right: The colour reflects the a-posteriori probability $P(\omega_1|\mathbf{x})$. In both images the $n_{\mathcal{V}} = 6$ prototypes are displayed as solid black diamonds. Discontinuities at the contact point of the voronoi cells can be observed by means of the vertices of the class border and the rapidly changing colouring in certain regions.

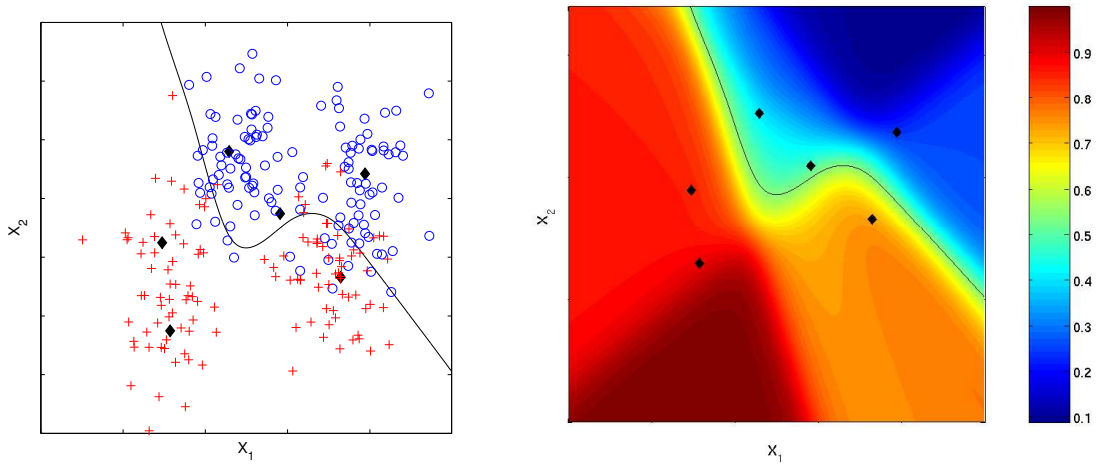


Figure 4.8: Same images as in Fig. 4.7, but calculated with a LSM operated in the soft-gating mode with $\phi = 1$. The $y_1(\mathbf{x}) = y_2(\mathbf{x})$ isoline (black) and the colour coding of $P(\omega_1|\mathbf{x})$ now varies smoothly and exposes no discontinuities at the contact points of two voronoi cells.

examples $\mathbf{x}_i, \mathbf{x}_j$ are close in the $\mathcal{X}$, the corresponding outcome $\mathbf{y}_i(\mathbf{x}), \mathbf{y}_j(\mathbf{x})$ can vary significantly if each example is located in a different voronoi cell, i.e. is evaluated by a different local expert.

In order to avoid such discontinuities, the LSM can also be operated in a *soft-gating mode*: Unseen examples are evaluated by all $n_{\mathcal{V}}$ local experts and the VQ layer acts as a soft-gating network which combines the individual outcomes $\mathbf{y}_l(\mathbf{x})$ of the $l = 1, \ldots, n_{\mathcal{V}}$ local experts, each

weighted by a coefficient $z_l(\mathbf{x})$, to the final result $\mathbf{y}(\mathbf{x})$ calculated by

$$\mathbf{y}(\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \sum_{l=1}^{n_\mathcal{V}} z_l(\mathbf{x}) \mathbf{y}_l(\mathbf{x}).$$

The normalisation coefficient

$$Z(\mathbf{x}) = \frac{1}{\sum_{m=1}^{n_\mathcal{V}} z_m(\mathbf{x})}$$

assures that the components of $\mathbf{y}(\mathbf{x})$ accumulate to one. The impact of l-th local expert on the final outcome is determined by a nonlinear function of the Euclidean distance $d(\mathbf{w}_l, \mathbf{x}) = \|\mathbf{w}_l - \mathbf{x}\|_2$ between the associated prototype $\mathbf{w}_l$ and the input $\mathbf{x}$ and is calculated by

$$z_l(\mathbf{x}) = \exp\Big(-\frac{d(\mathbf{w}_l, \mathbf{x})}{\phi}\Big).$$

The parameter ϕ controls the decay for an increasing distance $d(\mathbf{w}_l, \mathbf{x})$. For $\phi \to 0$, the outcome of the LSM operated in the soft-gating mode becomes equivalent to that of the LSM operated in the hard-gating mode.

The data of figure 4.7 but processed with a soft-gating LSM ($\phi = 1$) can be observed in figure 4.8. The influences of the smoothing properties become obvious if the course of the class border (black line) is investigated. In contrast to the solution of the hard-gating LSM, the class border smoothly varies in the considered region of the data space.

4.5 Assessment of Classification Performance

In the previous sections, different classification algorithms have been introduced which will be the fundamental components of the *computer aided diagnosis* (CAD) methods proposed in the following chapters. Each supervised classification algorithm comes with a certain number of hyperparameters which have to be adjusted manually by the operator to reasonable values. The choice of different algorithms as well as the variation of the algorithm's hyperparameters leads to classification models with varying characteristics and the operator has to decide which one is suitable for a given task. A second aspect, which is especially important for CAD systems in medical data domains, is the comparison of the performance of a CAD system with a *ground truth* as derived from an established diagnosis procedure such as a manual assessment of a human expert or the outcome of a laboratory test. Since this ground truth can also be interpreted as the outcome of a 'classifier', unique schemes can be derived for the comparison of classification models with a *reference model*, which can either be an accepted ground truth or a second classification model.

In the following, two schemes for assessing classifier performance will be introduced. The first scheme is based on the evaluation of the *confusion matrix* and thereby assesses the concordance of class labels assigned by two classifiers. The second scheme utilises the *receiver operating characteristics* (ROC) analysis for the assessment of classification models. In contrast to the confusion matrix, the ROC analysis does not only assess the accuracy of the assigned class label,

but also evaluates the discriminatory power of the assigned confidence value inducing a ranking on a set of classified examples. For both schemes, different indices are derived which summarise the performance of a certain classification model as a single scalar value.

4.5.1 Confusion Matrix Based Model Assessment

Most indices used for measuring the performance of classification models are based on the information provided by the *confusion matrix*. Assuming the classification model f as well as the reference model r attribute each example $\mathbf{x}_i \in \Gamma, i = 1, \ldots, N$ as a member of one of the n_Ω classes $\omega_k, k = 1, \ldots, n_\Omega$. In this case, the confusion matrix $\mathbf{C}$ is a $n_\Omega \times n_\Omega$ matrix with entries $\mathbf{C}_{kl}$ reflecting the number of examples designated as a member of ω_l by the reference model and as a member of ω_k by the investigated classification model. Thus, the diagonal entries $\mathbf{C}_{kk}$ represent consistently labelled examples, whereas the off-diagonal elements represent examples which are labelled differently by the two models. While the confusion matrix provides a detailed error statistic, different classification models are commonly compared by summarising this statistic by a suitable index. The most frequently used index is the *accuracy*

$$\mathsf{Acc} = \frac{\mathsf{trace}(\mathbf{C})}{N} \tag{4.43}$$

measuring the total fraction of consistently labelled examples for a fixed test set of N examples. Nevertheless, accuracy is an inappropriate measure if the a-priori probabilities $P(\omega_k)$ of the different classes differ significantly. Suppose 95% of the examples in Γ belong to ω_1, a classifier which attributes each presented example as a member of ω_1 yields a high accuracy of 0.95 although the classifier did not capture any information from the training data. In this situation, a better performance measure is the *balanced accuracy*

$$\mathsf{Acc}_{\mathsf{Bal}} = \frac{1}{n_\Omega} \sum_{k=1}^{n_\Omega} \frac{\mathbf{C}_{kk}}{N_{\omega_k}} \tag{4.44}$$

with N_{ω_k} being the number of examples $\mathbf{x}_i \in \Gamma$ which are members of class ω_k. The balanced accuracy measures the average class-specific accuracy. Alternatively, the chance-corrected concordance of two classification models can be measured using the κ-*coefficient* [Fleiss, 1982]

$$\kappa = \frac{N \sum_{k=1}^{n_\Omega} \mathbf{C}_{kk} - \sum_{k=1}^{n_\Omega} \mathbf{C}_{k\cdot}\mathbf{C}_{\cdot k}}{N^2 - \sum_{k=1}^{n_\Omega} \mathbf{C}_{k\cdot}\mathbf{C}_{\cdot k}} \tag{4.45}$$

with $\mathbf{C}_{k\cdot}, \mathbf{C}_{\cdot k}$ being the sum of entries in the k-th row or in the k-th column.

Another argument against employing accuracy as a performance measure is that different types of classification errors $\mathbf{C}_{kl}, k \neq l$ are recorded as being equally important. In particular in medical applications, certain types of misclassification cause different costs. For instance, attributing a diseased patient as healthy might by a fatal error, while a healthy patients attributed as diseased can be regarded as being a less serious error since this error can be corrected by future tests. In order to take into account the different costs as caused by different types of misclassification, the $n_\Omega \times n_\Omega$ *cost matrix* or *loss matrix* $\mathbf{L}$ with $L_{kl} \geq 0$ assigns a certain loss to each kind of error.

The *costs* of a certain classification model are calculated by

$$\text{Cost} = \sum_{k,l=1}^{n_\Omega} \mathbf{C}_{kl}\mathbf{L}_{kl}. \tag{4.46}$$

4.5.2 Receiver Operating Characteristics

Receiver operating characteristics (ROC) analysis [Swets and Pickett, 1982] is a technique for visualising and analysing the behaviour of decision making systems and has been the object of investigations in a large number of manuscripts examining different aspects of ROC analysis (see e.g. [Zou, 2002]). Whereas ROC analysis is a well established technique in medical decision making [Hanley, 1989, Obuchowski, 2003], it has recently been used with greater frequency for comparing artificial learning algorithms [Bradley, 1997, Flach, 2003, Provost and Fawcett, 2001].

The standard ROC analysis only considers binary problems. In order to be conform with the ROC literature, the classes are called positive ω_+ and negative ω_- indicating presence or absence of a certain attribute. Assuming that the examples of a data set Γ are classified by a binary classification model, which assigns a continuous score $f_{\boldsymbol{\theta}}(\mathbf{x}_i)$ to each example $\mathbf{x}_i \in \Gamma$. The score reflects the confidence that the example belongs to ω_+. For a comparison of the outcome with a reference model or ground truth by means of a confusion matrix $\mathbf{C}$, a threshold t_f has to be applied to the scores attributing the example $\mathbf{x}_i$ as an example of the positive class if $f_{\boldsymbol{\theta}}(\mathbf{x}_i) > t_f$ and, otherwise, as an example of the negative class. The 2×2 confusion matrix summarising the error statistics consists of the diagonal elements[2]:

True-Positives (TP) : Number of positive examples which are attributed consistently by both models,

True-Negatives (TN) : Number of negative examples which are attributed consistently by both models.

The off-diagonal elements reflect inconsistently labelled examples are called:

False-Positives (FP) : Number of true negative examples which are attributed as positive by the classification model,

False-Negatives (FN) : Number of true positive examples which are attributed as negatives by the classification model.

Based on these four quantities, further indices can be utilised for comparing different classification models:

Sensitivity (SE) or True-Positive-Rate (TPR) : Ratio of the number of positive examples consistently labelled by the classification model and the total number of positives

$$\text{SE} = \frac{\text{TP}}{\text{TP}+\text{FN}},$$

[2] To conform with the literature, one model is considered as *classification model* which is to be assessed, while the second model is considered as ground truth reflecting the *true* class label of an example.

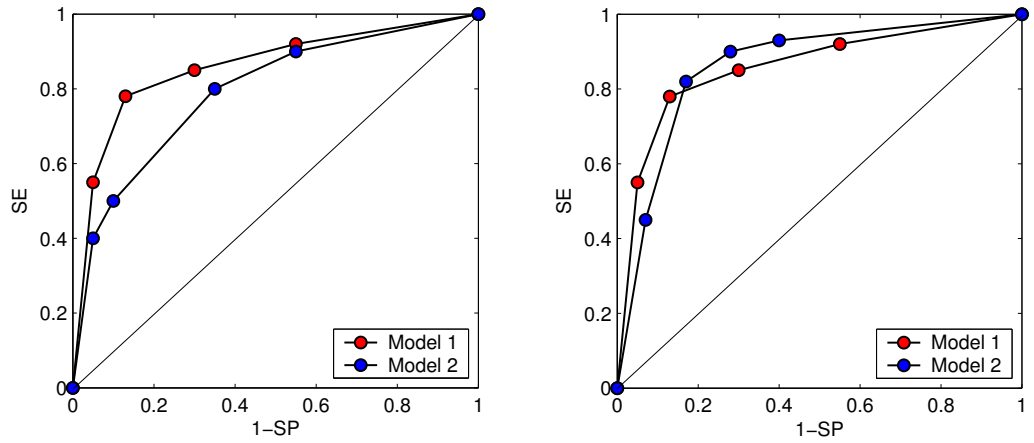


Figure 4.9: ROC curves for different classification models. The performance of a model randomly attributing the examples as positive or negative is visualised by the black diagonal line. For each curve the operating points of six different values of t_f are plotted. Left: The curves illustrate the superior performance of Model 1. The plot indicates an increased SE of Model 1 for the entire range of SP. Therefore, Model-1 is a *dominating model* [Provost et al., 1998] which is at least as good as all other regarded models. Right: The ROC curve for Model 1 shows a superior SE for $(1 - \text{SP}) < 0.18$, whereas Model 2 outperforms Model-1 in terms of SE for $(1 - \text{SP}) \geq 0.18$.

Specificity (SP) : Ratio of the number of negative examples consistently labelled by the classification model and the total number of negatives

$$\text{SP} = \frac{\text{TN}}{\text{TN}+\text{FP}}.$$

False-Positive-Rate (FPR) : Ratio of the number of false positives and the total number of negatives

$$\text{FPR} = \frac{\text{FP}}{\text{FP}+\text{TN}} = 1 - \frac{\text{TN}}{\text{TN}+\text{FP}} = 1 - \text{SP}.$$

Positive-Predictive-Value (PPV) : Ratio of the number of positive examples consistently labelled by the classification model and the total number of examples classified as positive

$$\text{PPV} = \frac{\text{TP}}{\text{TP}+\text{FP}}.$$

.

ROC Curves

The SE and SP for a certain t_f only provide limited information about the classification model because both quantities depend on the choice of t_f. Increasing t_f typically raises the SE, but simultaneously lowers the SP. Thus, t_f controls the trade-off between SE and SP and different values of t_f define different *operating points* $(\text{SE}, \text{SP})_{t_f}$ in the ROC space $\mathcal{R} = [0; 1]^2$. The ROC

space is commonly visualised by a two-dimensional plot with SE on the y-axis and (1-SP) on the x-axis. Less formally, an operating point suggests a superior performance compared to a second one if it is located closer to the upper right corner of $\mathcal{R}$, i.e. (1-SP,SE)=(1,1).

If all possible operating points for a data set Γ as evaluated by a classification model are plotted and connected by a line, the tradeoff-characteristics of a model can be displayed by the *empirical ROC curve* in $\mathcal{R}$ running from $(1-\text{SP}, \text{SE}) = (0,0)$ to $(1-\text{SP}, \text{SE}) = (1,1)$. The ROC curve of a classification model randomly attributing examples as positive or negative is a straight line from $(1-\text{SP}, \text{SE}) = (0,0)$ to $(1-\text{SP}, \text{SE}) = (1,1)$. Each ROC curve above this diagonal reflects a certain discriminatory power for the corresponding mode.

Figure 4.9 illustrates the ROC curves for two pairs of classification models. The course of the ROC curves provides detailed information about the SE/SP trade-off characteristics of each classification model which is in particular important if there is no *dominating model* [Provost et al., 1998] such as Model-1 in Fig. 4.9, left. However, different models can be superior for different regions of $\mathcal{R}$ (Fig. 4.9,right).

An appealing property of the ROC curve is that it provides an assessment which is insensitive to unbalanced or changing class distributions [Provost et al., 1998], because each of the measure SE and SP considers either the positive (TP,FN) or negative examples (TN,FP). Thus, the ROC curve does not change for data sets Γ with varying ratios of the number of positive and negative examples [Fawcett, 2003], which is common in real-world applications.

Averaging ROC Curves

It is common practice to assess the performance of a classification model by means of a repeated application to pairs of training and test sets in order to estimate the average performance. Such pairs of training and test sets can either be different random partitions of the entire data or the more systematically splitting of the training data in a cross-validation process. Nevertheless, each of the n_p pairs of training and test sets leads to a single ROC curve and all n_p curves have either to be simultaneously considered or suitably combined for the assessment of the model. There are basically two techniques for combining ROC curves: *Pooling* and *averaging*. Using *pooling* [Swets and Pickett, 1982], the examples of all n_p test sets together with the corresponding outcome of the classification model are merged into one large set and a single ROC curve is calculated for this set. In *averaging* [Fawcett, 2003, Provost et al., 1998], an averaged ROC curve is calculated by sampling the n_p ROC curves at a certain number of (1-SP) points (*vertical averaging*) or at a certain number of thresholds t_f (*threshold averaging*). The averaging technique has the advantage that the variance of sampled values can be considered and easily displayed as a *confidence band* around the averaged ROC curve.

Selection of Optimal Thresholds

For practical applications, the classification model and a reasonable operating point, i.e. a specific value of threshold t_f, has to be chosen under consideration of the demands of the application. Such a requirement can e.g. be a fixed value or range of SE that has to be met by the model. If the costs c_{FP} and c_{FN} for a FP or FN example can be quantified, the optimal classifier can be chosen by plotting the *iso-performance lines* [Provost and Fawcett, 1997] in $\mathcal{R}$. The slope of the

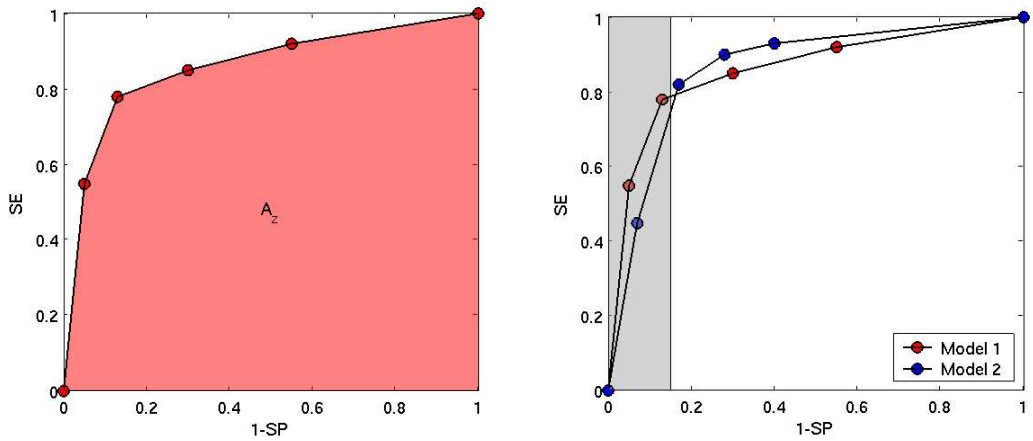


Figure 4.10: Left: The area under to ROC curve A_z (red) as a measure summarising the course of the curve. Right: Depending on practical considerations, the area under the ROC curve in a certain part of the ROC space, e.g. the range 1-SP=[0;0.15] (gray area) in the figure, provides a more meaningful assessment of two models.

line is calculated by

$$s(c_{\text{FP}}, c_{\text{FN}}) = \frac{P(\omega_+)c_{\text{FP}}}{P(\omega_-)c_{\text{FN}}} \tag{4.47}$$

with $P(\omega_+)$ and $P(\omega)$ being the a-priori probability of the classes ω_+ and ω_-, respectively. Each classifier with an operating point lying on this line causes the same costs under the given circumstances. The optimal operating point for a certain model is given by that operating point which is closest to that point where the iso-performance line is tangent to the model's ROC curve. In the case of unknown or changing costs, Provost and Fawcett, 2001 propose the application of the *ROC Convex Hull* technique which combines all classification models which are dominant in limited regions of $\mathcal{R}$. For each operating point, this *hybrid classification model* will perform at least as good as any of the individual models considered.

The Area-Under-the-ROC-Curve Index

ROC curves are two-dimensional depictions of classification models. In order to summarise each of these curves by a single index, the *area-under-the-ROC-curve* (A_z) index is frequently calculated (see Fig. 4.10, left) [Swets and Pickett, 1982, Hanley and McNeil, 1982, Bradley, 1997]. The value of A_z for a certain curve can be estimated by either a parametric or a nonparametric model. In the former case, a bivariate normal distribution is fitted to the outcome $f_{\boldsymbol{\theta}}(\mathbf{x}_i), \mathbf{x}_i \in \Gamma$, which implicitly assumes that the class-specific outcomes follow a normal distribution. In the latter case, A_z is calculated by the *trapezoidal rule* and is equivalent to the outcome of the *Wilcoxon-Mann-Whitney* statistic [Hanley and McNeil, 1982] and closely related to the *Gini* index [Breimann et al., 1998]. The parametric approach tends to underestimate the true area, but this bias is commonly negligible for a continuous model output $f_{\boldsymbol{\theta}}(\mathbf{x})$. Hanley and McNeil, 1982 also derived

the formula

$$\epsilon(A_z) = \sqrt{\frac{A_z(1-A_z) + (m_{\omega_+} - 1)(Q_1 - A_z^2) + (m_{\omega_-} - 1)(Q_2 - A_z^2)}{m_{\omega_+} m_{\omega_-}}} \tag{4.48}$$

for the *standard error* $\epsilon(A_z)$ with $Q_1 = A_z/(2 - A_z)$ and $Q_1 = 2A_z^2/(1 + A_z)$.

Several interpretations exist for the A_z value such as the average SP for all possible values of SE [Metz, 1986] or the probability that a randomly selected positive example $\mathbf{x}_+$ leads to an outcome $f_{\boldsymbol{\theta}}(\mathbf{x}_+)$ which is larger than the outcome $f_{\boldsymbol{\theta}}(\mathbf{x}_-)$ for a randomly selected negative example $\mathbf{x}_-$ [Hanley and McNeil, 1982].

The Partial-Area-Under-the-ROC-Curve Index

In a case where practical considerations like limited range of admissible SP induces constraints on the location of a suitable operating point, the A_z indices can provide misleadingly assessments of two models if the corresponding curves cross. Both ROC curves in Fig. 4.10 (right) yield a comparable A_z value which suggest a similar performance of both models. Nevertheless, if operating points in the range of e.g. $0 \leq (1 - \text{SP}) \leq 0.15$ are practically relevant, Model 1 is preferable. In order to obtain are more meaningful index which considers such practical constraints, McClish, 1989 and Jiang et al., 1996 suggest to determine the *partial-area-under-the-ROC-curve* (A_x) with x denoting a particular limitation. Such limitations can either be a certain range of SE (e.g. $A_{(\text{SE} \geq 0.9)}$) or a certain range of SP (e.g. $A_{(0 \leq (1-\text{SP}) \leq 0.15)}$). Furthermore, Jiang et al., 1996 propose the *standardised partial area* index $0 \leq A_x \leq 1$ which is the partial area normalised by the largest possible area value.

Comparing A_z Indices of Different Models

When comparing two models M_1 and M_2 by means of the corresponding indices A_z^{M1} and A_z^{M2} the question arises, whether the difference $A_z^{M1} - A_z^{M2}$ is statistically significant. If the A_z value for M1 and M2 is calculated using two disjoint test sets, the critical ratio is

$$z = \frac{A_z^{M1} - A_z^{M2}}{\epsilon(A_z^{M1} - A_z^{M2})} \tag{4.49}$$

with $\epsilon(A_z^{M1} - A_z^{M2})$ as the estimated standard error for the difference between the two areas [Hanley and McNeil, 1983] as calculated by

$$\epsilon(A_z^{M1} - A_z^{M2}) = \sqrt{\epsilon(A_z^{M1})^2 + \epsilon(A_z^{M1})^2}. \tag{4.50}$$

The quantity z is then compared to tables of normal distribution to test for significance. Nevertheless, if A_z^{M1} and A_z^{M2} are calculated on the same test set, both values are likely to be correlated. Hanley and McNeil, 1983 considers this correlation by incorporating the term $2r\epsilon(A_z^{M1})\epsilon(A_z^{M1})$ in (4.50)

$$\epsilon(A_z^{M1} - A_z^{M2}) = \sqrt{\epsilon(A_z^{M1})^2 + \epsilon(A_z^{M1})^2 - 2r\epsilon(A_z^{M1})\epsilon(A_z^{M1})} \tag{4.51}$$

with a tabulated quantity r depending on the average area $(A_z^{M1} + A_z^{M2})/2$ and the average of the class-specific correlation coefficient $(r_{\omega_+} + r_{\omega_-})/2$. An alternative approach for comparing two or more correlated ROC curves adopts the theory on *generalised U-statistics* for covariance matrix estimation [DeLong et al., 1988].

5 Lesion Detection

According to Heywang-Köbrunner and Beck, 1995, localisation of tissue suspected of being affected by pathophysiological disorders is the first step in DCE-MRI based breast cancer diagnosis. Temporal kinetic signals exhibiting suspicious signal characteristics may be signs of benign or malignant disorders of the underlying tissue and require for a more detailed examination by either manually assessing the courses of the corresponding kinetic signals or by applying more sophisticated techniques for analysis.

A standard procedure for localising suspicious masses used in clinical practice is the examination of DCE-MRI sequences by means of one or several subtraction images, each displaying the temporal gradient of intensity values between the precontrast image and one of the postcontrast images. Tissue regions exhibiting signal enhancements appear in subtraction images as clusters of voxels with high intensity values. Nevertheless, examination of the data by means of a single subtraction image can be suboptimal due to the heterogeneity of cancerous tissue. In particular carcinoma typically consist of a conglomerate of benign and malignant tissue. Whereas in malignant tissue the concentration of *contrast agent* (CA) molecules rapidly increases to a peak in the very early postcontrast period and decreases afterwards, benign tissue commonly exhibits continuously increasing signals which reach the peak of CA concentration in the late postcontrast period. Hence, voxels of malignant tissue are likely to be displayed with high intensity in subtraction images based on the very early postcontrast images, while subtraction images based on the late postcontrast images are likely to be more suitable for the detection of benign tissue. Next to subtraction images displaying the temporal gradient caused by the uptake of CA concentration between a pre- and a postcontrast image, the temporal gradient caused by the wash-out of CA molecules in malignant tissue can be displayed by means of subtraction images using one late and one early postcontrast image. In consequence, the observer has to correlate the content of several images by switching between or by side-by-side display of subtraction images based on different postcontrast images. In either case, a large number of images has to be examined for each case, since it is still common practice to display the three-dimensional image data as a sequence of two-dimensional image slices.

Subtraction images provide the observer access to the DCE-MRI data, which weights the spatial dimensions of the data higher than the temporal dimension. A single subtraction image retains the entire spatial information of the DCE-MRI data as given by the coordinates of voxels but only considers a fragment of the temporal content. In contrast, the data-driven approach to DCE-MRI data analysis as proposed in this chapter temporarily ignores the spatial information of the data and focuses on the evaluation of the temporal aspects. Supervised classification algorithms are applied to distinguish voxels of normal and voxels of suspicious tissue solely by means of the associated temporal kinetic signals. Instead of defining explicit models of the temporal kinetic signals as measured for the two tissue types, implicit models are derived from the DCE-MRI data itself. During an adaptation process, an *artificial neural network* (ANN) correlates the signal

information of a small number of training cases with an expert label, marking the position of those voxels, which gave rise to the radiologist's diagnosis. As described below, such a label is not optimal for the purpose of training an ANN but reflects the typical amount of class information which can be acquired during a common diagnosis process which is geared to the medical problem of cancer diagnosis and not to the demands of an ANN application.

The remainder of this chapter is structured as follows. In section 5.1 the setup for lesion detection with ANNs is described including data preprocessing, sampling of training data and feature computation. Detection results for two different groups of cases are presented in section 5.2 and discussed in the final section. The objectives of this chapter are

- to describe a data-driven approach to lesion detection based on supervised artificial neural networks, which avoids the requirement of an explicit definition of mathematical models for the temporal kinetic signals of normal and suspicious tissue,
- to evaluate whether and how well implicit models can be trained by correlating signal information given by a small number of DCE-MRI sequences with additional diagnostic information acquired from a clinical diagnosis process,
- to describe how the trained ANN can be utilised to generate a single image considering the entire temporal information of the DCE-MRI data and which displays lesions with high contrast to normal tissue reducing the requirements of time-consuming and subjective manual optimisations of the data presentation.

5.1 Detection of Lesions with ANNs

Localisation and delineation of the extent of lesions are regarded as a binary classification problem in which voxels representing normal tissue have to be distinguished from voxels of malignant or benign tissue. A sketch of the lesion detection setup is given in figure 5.1. Fundamental ideas and details of the setup are discussed in the following.

Data Distribution of DCE-MRI Feature Vectors

A DCE-MRI sequence is interpreted as a set of temporal kinetic signals $\{\mathbf{s_p}\}$ on which a spatial order is induced by the coordinates $\{\mathbf{p}\}$ of the associated voxels. If this spatial order is ignored, the temporal kinetic signals of one or several DCE-MRI sequences form a distribution of points in a signal space $\mathcal{S}$. Each point in $\mathcal{S}$ represents the temporal course of the signal intensity of one voxel. Since the signal space only considers the temporal information of the DCE-MRI data, similar temporal kinetic signals are likely to be close to each other in the signal space, irrespective of the spatial location of the corresponding voxels in the three-dimensional image volume. Therewith, temporal kinetic signals of different tissue types exhibiting tissue-specific signal characteristics are likely to be concentrated in certain subregions of the signal space. The approach proposed in the following aims to identify such class-specific subregions of the signal space by analysing the data distribution of signals of normal and suspicious tissue using supervised artificial neural networks. In order to obtain a representation of the data which is even more suitable for the evaluation by supervised classification algorithms, the data are mapped by a transformation T to

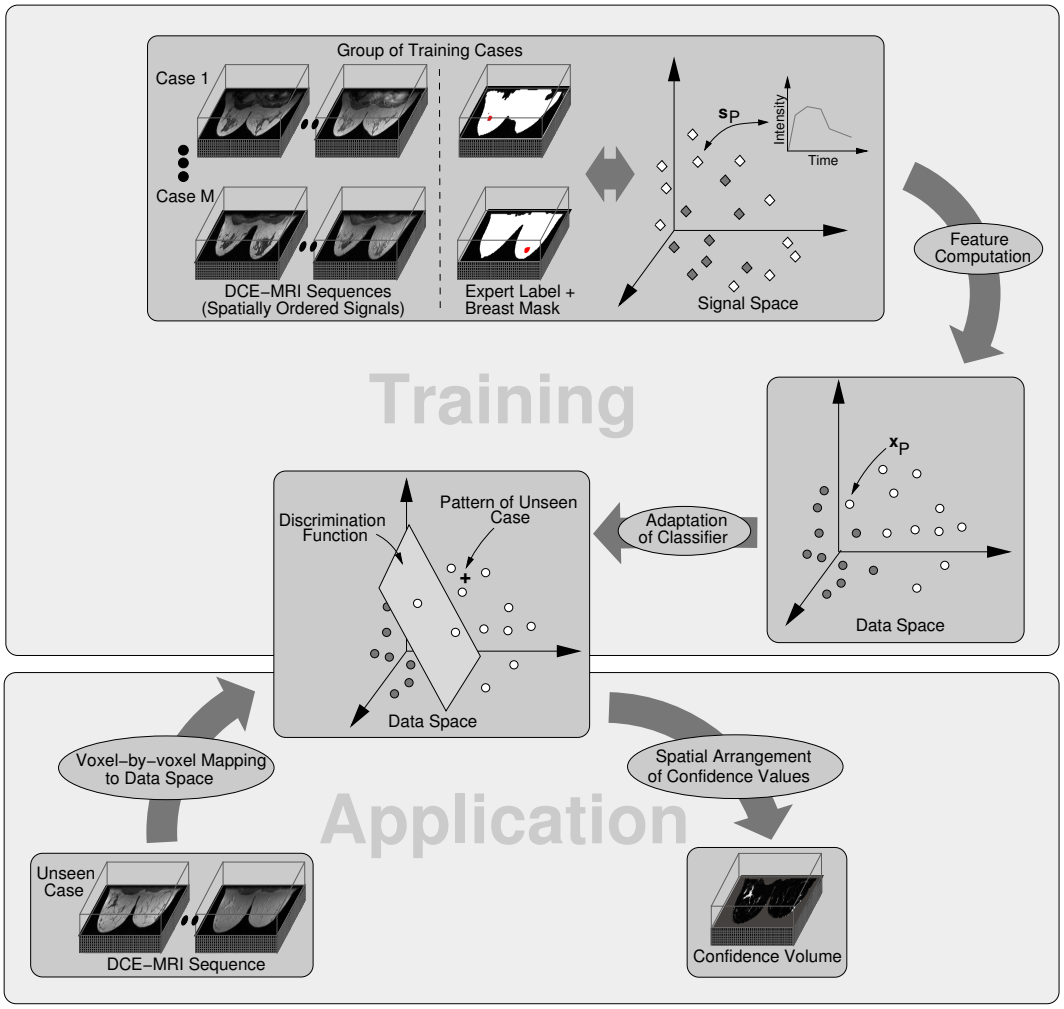


Figure 5.1: Overview of the setup for detecting suspicious lesions using adaptive classification algorithms. If the spatial order given by the voxel coordinates is ignored, the temporal kinetic signals of several training cases can be considered as a data distribution in a signal space $\mathcal{S}$. Using the expert label and a binary breast mask generated during a preprocessing step, signals of the training cases can be labelled as examples for either normal or suspicious breast tissue. These training examples are mapped from $\mathcal{S}$ to a data space $\mathcal{X}$, in which each example is described by a vector of suitable features $\mathbf{x} \in \mathcal{X}$. Using the training data, an adaptive classification algorithm is trained to distinguish examples from the two different tissue classes. Subsequently, the resulting discrimination function can be applied for a voxel-by-voxel evaluation of DCE-MRI sequences of unseen cases. To this end, a feature vector $\mathbf{x_p} \in \mathcal{X}$ is calculated for each temporal kinetic signal $\mathbf{s_p} \in \mathcal{S}$ and is evaluated by the trained predictor. The computed confidence value $y(\mathbf{x_p})$ is assigned to the voxel at the corresponding position $\mathbf{p}$ in a new three-dimensional *confidence volume*, which displays lesion tissue by voxels with high intensities.

a new data space $\mathcal{X}$. Two different groups of transformations are investigated. The first group are *voxel-based transformations* which map each temporal kinetic signal $\mathbf{s_p}$ to a feature vector

$\mathbf{x_p}$:

$$T : \mathcal{S} \to \mathcal{X}, \quad \mathbf{s} \mapsto \mathbf{x}. \tag{5.1}$$

This group of transformations also includes the identity function which allows to consider the signal space $\mathcal{S}$ as a variant of a data space $\mathcal{X}$. The second group are *region-based transformations*

$$T : \mathcal{S}^{n\mathbf{W}} \to \mathcal{X}, \quad \mathbf{W} \mapsto \mathbf{x}, \tag{5.2}$$

which map each $w_x \times w_y \times w_z$ neighbourhood or window $\mathbf{W_p}$ consisting of $n_{\mathbf{W}}$ voxels with centre at $\mathbf{p}$ to a feature vector $\mathbf{x_p}$ representing the variation of the region's texture over time. Both groups of transformations will be described in more detail in a later section.

Supervised Analysis of Feature Vectors

Instead of explicitly defining mathematical models for the feature vectors of different tissue classes using e.g. medical a-priori knowledge, which is difficult for feature vectors obtained from a transformation other than the identity function, implicit models are derived from the data itself. The feature vectors computed from the data of a small number of training cases are labelled according to the segmentation of the lesions by a radiologist and are subsequently utilised for the adaptation of a supervised classification algorithm. The classification algorithm is trained to distinguish the feature vectors of the two tissue classes and provides a discrimination function

$$C : \mathcal{X} \to [0; 1], \quad \mathbf{x} \mapsto y(\mathbf{x}),$$

which maps each pattern $\mathbf{x_p}$ to a single scalar value $y(\mathbf{x_p})$ indicating the confidence that the evaluated pattern was caused by suspicious tissue. Three different classifiers are investigated. The *support vector machine with linear kernel* (SVM-L) and the *linear discriminant analysis* (LDA) provide linear discrimination functions which divide the data space by a hyperplane into two half-spaces. The *support vector machine with Gaussian kernel* (SVM-G) distinguishes examples of the two classes by a more complex non-linear discrimination function.

Acquisition of Labelled Data

The label attributing each feature vector as a normal or suspicious example is obtained from a manual lesion segmentation acquired during a standard diagnosis process. A radiologist designated those subregions of the DCE-MRI data representing tissue masses which give rise to his final diagnosis. Even though such a label reflects a typical clinical proceeding which aims for an optimal diagnosis for each individual patient, it is suboptimal for the purpose of a data-driven evaluation of the DCE-MRI data for two reasons:

1. The label is incomplete because only those voxels are marked which directly give rise to the final diagnosis. For instance, if the DCE-MRI sequence of a patient exhibits a malignant disorder, the region designated by the radiologist commonly includes only those voxels displaying the carcinoma. In this case, voxels of additional benign disorders such as fibrocystic changes commonly remain unmarked since they have only minor impact on the final

diagnosis and the following treatment. Because of the lack of a label for the entire image volume, such unmarked disorders have to be considered as normal tissue in the data-driven analysis.

2. The label only provides an uncertain assessment of the informational content of the temporal kinetic signals. For diagnosing breast cancer, radiologists commonly correlate information obtained from different sources such as different imaging modalities (DCE-MRI, X-ray mammography, ultrasound) or physical examinination (palpation). Furthermore, morphological properties of larger conglomerates of voxels displaying a suspicious disorder are considered in addition to the temporal information provided by the kinetic signals. For instance the spatial heterogeneity of the signal enhancement or properties of the lesion margin are features of larger regions of tissue depicted by a conglomerate voxels and provide valuable information for deciding whether a tissue mass has to be considered for the diagnosis. Thus, the label assigned to a voxel does not necessarily reflect the outcome which would have been obtained if solely the temporal kinetic signals of the individual voxels had been assessed sequentially and independently from each other.

Nevertheless, the given label represents an amount of information which can be acquired during an evaluation process which is primarily geared to the demands of medical diagnosis. Even the creation of such incomplete labels for ten cases took about one working day of a radiologist and a computer scientist, who evaluated the corresponding subtraction images with a clinical standard software.

Evaluation of Unseen Cases

After adaptation of a classifier with the labelled data, unseen DCE-MRI sequences can be evaluated. The image data of an unseen case is mapped to the same data space as the training data. Subsequently, the trained classifier evaluates each feature vector $\mathbf{x_p}$ and returns a confidence value $y(\mathbf{x_p}) \in [0;1]$, indicating the probability that $\mathbf{x_p}$ was caused by suspicious tissue. Due to the well-defined range of the confidence values, each $y(\mathbf{x_p})$ can be interpreted as a gray level and is assigned to the voxel at the spatial position $\mathbf{p}$ in a new three-dimensional image, referred to as *confidence volume*. This volume enables the radiologist to localise lesions and to delineate their extent by examination of a single image, because tissue masses exhibiting suspicious kinetic signals are displayed as clusters of bright voxels.

In the remainder of this section, the preprocessing of the image data, the selection of training data and the calculation of feature vectors are described in more detail. The following section will then describe the application of the setup to two groups of cases. The first group, referred to as *Munich* group, consists of DCE-MRI sequences of six benign and six malignant cases recorded at the City Centre Hospital of the University of Munich. The second group, referred to as *MARIBS* group, consists of five malignant and three benign cases who are participants of the *MARIBS* breast cancer screening study.

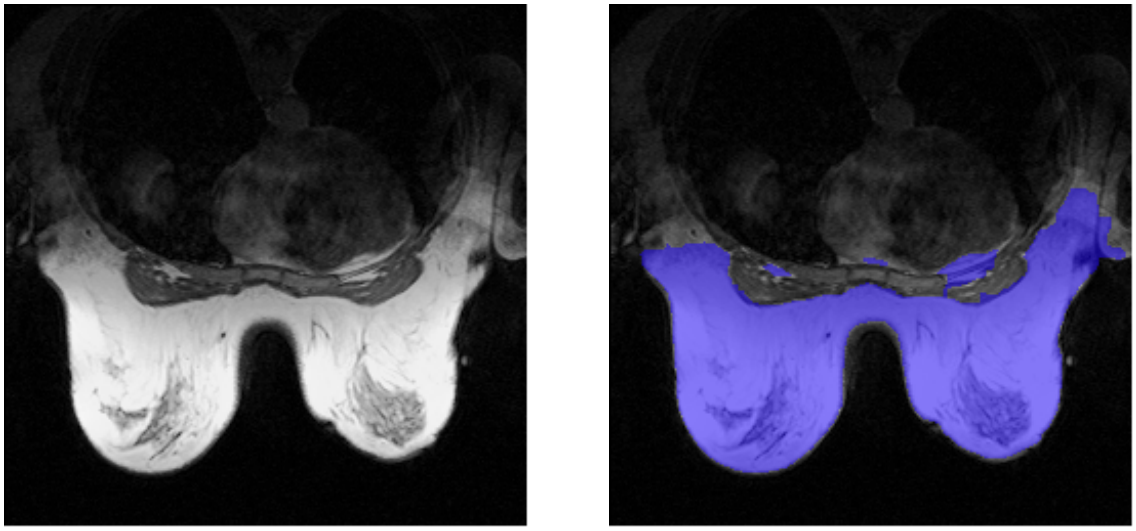


Figure 5.2: Left: Axial slice of a precontrast image volume recorded at the City Centre Hospital of the University of Munich. Gray values have been scaled to enhance the display of the thorax region. Right: Final breast mask (red) for the axial slice shown in the left image.

5.1.1 Preprocessing of Image Data

Depending on the protocol utilised for the acquisition of DCE-MRI sequences, the recorded images display different field-of-views, i.e. different subregions of the body. Most setups for MRI based examination of the breast allow for simultaneous imaging of both breasts. Thereby, the cuboid shaped field-of-view typically encloses a significant fraction of background such as non-body regions (air) or regions of the body which are outside the scope of investigation. Before applying the adaptive lesion detector, this background is separated from breast tissue to limit the random sampling of normal tissue signals used for adaptation of the classifiers to the region of the breast and to accelerate the evaluation of unseen cases.

Preprocessing of *Munich* Images

The DCE-MRI setup for breast examination employed at the City Centre Hospital of the University of Munich assumes patients to be placed in a prone position with the chest resting on the dedicated surface coil. The image volumes display 32 to 34 parallel axial slices with a mutual distance of $4mm$ recorded using a multi-slice protocol. Each slice displays a section of the breast and the entire thorax with a field-of-view $350mm \times 350mm$ and a matrix size of 256×256 regular shaped pixels. Therewith, each voxel of the image represents a $1.37mm \times 1.37mm \times 4.0mm$ tissue volume.

Figure 5.2 (left) shows an example for an axial slice of a precontrast image of one case from the *Munich* group. Good signal-to-noise ratios are obtained for the tissue adjacent to the surface coil surrounding the breasts. The homogeneity and magnitude of the magnetic field decreases with the distance to the surface coil causing tissue deeper in the body to be displayed with signal intensities which are lower than those of the tissue at the body surface next to the coil. As

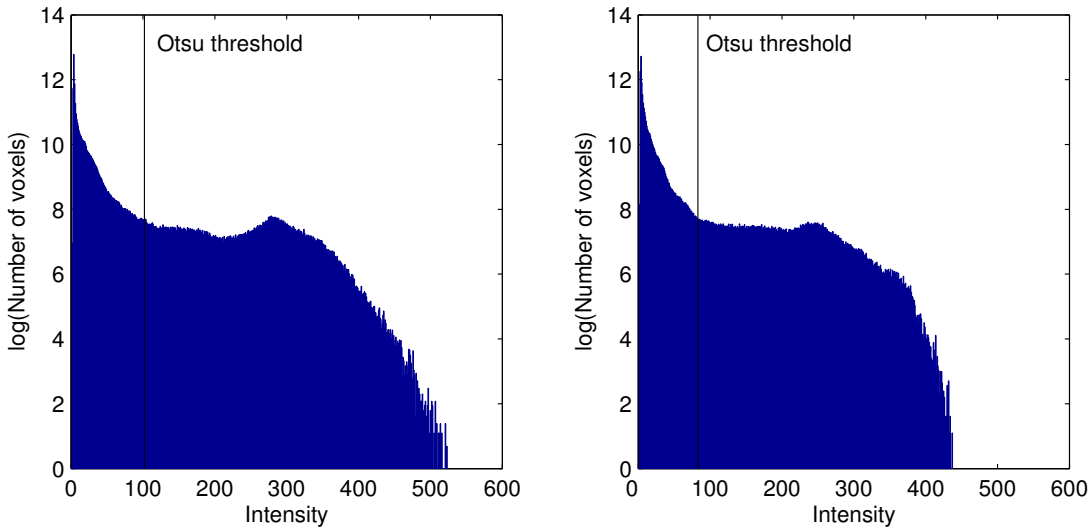


Figure 5.3: Intensity histogram of the median filtered precontrast image of case M6 (left) and B6 (right). Both histograms show a clear peak at low intensity values predominately caused by the large fraction of background. The black line illustrates the thresholds computed using the algorithm proposed by Otsu, 1979.

a consequence, the precontrast image displays the breast tissue with higher intensities than the tissue of the thorax. In particular the heart, which appears with high intensities in the postcontrast images due to the presence of large amounts of contrast agent molecules and blood in the heart cavities, is displayed by voxels of the same low intensity as the surrounding organs.

For separating the breast region from background, a binary mask for the breast is calculated using an adaptive threshold algorithm. First, the precontrast image is filtered using a 3×3×3 median filter. Afterwards, the adaptive threshold algorithm proposed by Otsu, 1979 is applied to the histogram of the filtered precontrast image (Fig. 5.3). The algorithm calculates a threshold value dividing the histogram into two classes in such a way that the ratio of inter- and intra-class variance is maximised. The major fraction of voxels corresponding to breast tissue expose intensities above the computed threshold. Binarisation of the filtered precontrast image using the determined threshold yields a binary mask covering the entire breast. Small gaps may appear in the centre of each breast where the fibro-glandular discs are located, which are typically displayed with low gray values. These gaps are closed by the application of a 5×5×5 morphological closing operator, i.e. a successive application of a morphological dilatation and erosion operator, to the binary breast mask. Figure 5.2 (right) visualises an axial slice of a precontrast image with the corresponding breast mask, which generously covers the breast tissue. The final binary mask, referred to as *breast mask*, determined for case m is formally described by the set $\mathrm{P}_m^{\mathrm{Breast}}$ containing all positions $\mathbf{p}$ of kinetic patterns $\mathbf{x_p}$ which are to be considered during the application of the learning algorithm.

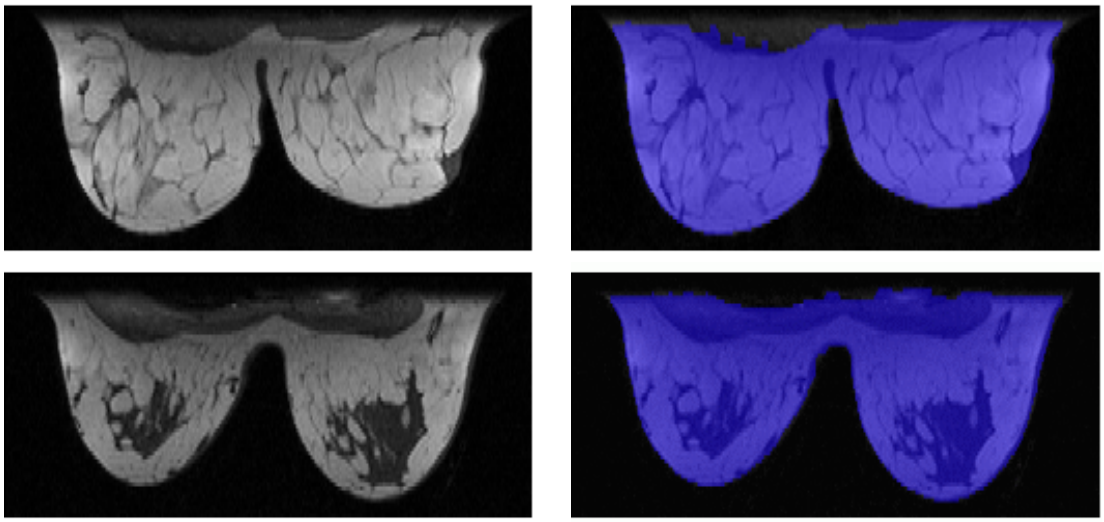


Figure 5.4: Image slices in coronal orientation of the precontrast images (left) and the corresponding breast masks (right) of two cases of the *MARIBS* group.

Preprocessing of *MARIBS* Images

Image volumes acquired from participants of the MARIBS screening study consist of $256 \times 128 \times 64$ voxels. The voxels have a square shape of $1.33mm \times 1.33mm$ in the coronal plane and an extent of $2.5mm$ in the perpendicular direction. In contrast to the *Munich* images, the field-of-view is limited to both breast and displays only a small fraction of the thorax. Thus, the aim of the preprocessing step is to separate the breast from the background which, for this pool, mainly consists of voxels displaying air.

Similar to the preprocessing of the *Munich* images, a binary breast mask is calculated based on the application of the adaptive threshold algorithm proposed by Otsu, 1979. To this end, the three-dimensional image volumes of each sequence are filtered using a three-dimensional median filter of size 3×3×3. Subsequently, a single scalar value is determined for each voxel by adding the temporal intensity gradients between two subsequent time points to the intensity value in the precontrast image:

$$s(\mathbf{x_p}) = x_{\mathbf{p}1} + \sum_{i=2}^{n_t} |x_{\mathbf{p}i} - x_{\mathbf{p}i-1}|.$$

From the bimodal shaped histogram of the new image, a threshold is automatically determined using the adaptive threshold algorithm. Application of a 5×5×5 morphological closing operator to the binarised image closes small gaps in the regions of the glandular discs which are displayed larger than in the *Munich* images due to a higher spatial resolution of the images. The final *breast mask* completely covers the regions of the breast and, for several cases, small parts of the heart (Fig.5.4).

5.1.2 Training Data Selection

For the adaptation of the binary classification algorithm, a training set $\Gamma_{\text{Train}} = \{(\mathbf{x}, \mathbf{y})_i\},\ i = 1, \ldots, N$ is selected consisting of tuples of an input vector $\mathbf{x}_i$, which is a point in an n_{in}-dimensional data-space $\mathcal{X}$, and a class label y_i. The class label attributes the example as a member of the class of normal tissue signals or as a member of the class of suspicious, i.e. malignant or benign, tissue signals. Examples for the latter class are sampled using the manual lesion segmentation of the radiologist. The outcome of the segmentation is a set $\mathsf{P}_m^{\text{Lesion}}$ containing the positions $\mathbf{p}$ of voxels displaying the diagnosed disorder of case m. Consequently, the set $\Gamma^{\text{Suspicious}}$ of training examples for the class of suspicious signals contains all patterns which are located at positions designated by the radiologist:

$$\Gamma^{\text{Suspicious}} = \{(\mathbf{x}, y)_{\mathbf{p}} | \mathbf{p} \in \mathsf{P}_m^{\text{Lesion}}\},\quad m = 1, \ldots, M. \tag{5.3}$$

Examples representing kinetic signals caused by normal tissue are randomly sampled from positions inside the breast mask, excluding the voxels marked as lesion:

$$\Gamma^{\text{Normal}} = \{(\mathbf{x}, y)_{\mathbf{p}} | \mathbf{p} \notin \mathsf{P}_m^{\text{Lesion}} \wedge \mathbf{p} \in \mathsf{P}_m^{\text{Breast}}\},\quad m = 1, \ldots, M. \tag{5.4}$$

The number of normal tissue signals commonly exceeds the number of suspicious tissue signals by far. To limit computational expenses and to avoid a decrease in classification performance caused by unbalanced training sets, only a subset of Γ^{Normal} is used for training. The exact composition of the training set varies for the different classification algorithms and is described in more detail in section 5.1.4.

5.1.3 Feature Description

The input patterns $\mathbf{x}_{\mathbf{p}} \in \mathcal{X}$ consist of features describing the state of the local tissue. For the given task, two groups of features are investigated. The first group consists of features $\mathbf{x}_{\mathbf{p}}$ computed solely from the temporal kinetic signal $\mathbf{s}_{\mathbf{p}}$ as measured for a voxel $\mathbf{p}$. The second group consists of features derived from a small neighbourhood surrounding the centre voxel $\mathbf{p}$ and describe the temporal variation of textural characteristics of the voxel's neighbourhood.

Voxel-Based Temporal Kinetic Features

Voxel-based temporal kinetic features $\mathbf{x}_{\mathbf{p}}$ solely utilise information as derived from the measured temporal kinetic signal $\mathbf{s}_{\mathbf{p}}$ describing the course of the local CA concentration in the voxel element at position $\mathbf{p}$. The different types of voxel-based temporal kinetic features stem from the two assumptions that signals of suspicious tissue predominantly differ from signals of normal tissue because of their temporal dynamics, i.e. the qualitative signal course, or their signal amplitudes at different points of time reflecting the amount of change. For the task of suspicious lesion detection, the following types of feature vectors are investigated:

- The *raw*-feature $\mathbf{x}_{\mathbf{p}} = \mathbf{s}_{\mathbf{p}} \in \mathcal{X} = \mathbb{R}^{n_t}$ describing the temporal course of the intensity value of the voxel $\mathbf{p}$.

- The *preratio*-feature

$$\mathbf{x_p} = (s_{\mathbf{p}2}/s_{\mathbf{p}1}, \ldots, s_{\mathbf{p}n_t}/s_{\mathbf{p}1}) \in \mathcal{X} = \mathbb{R}^{n_t-1} \tag{5.5}$$

reflecting the ratio of the intensity values as measured in the postcontrast images and the intensity value as measured in the precontrast images. This feature vector qualitatively describes the temporal dynamic of the signal and provides no information about the amount of signal enhancement at different points of time. An increase of the signal intensity from 10 to 20 is considered to be equal to a signal uptake from 100 to 200. A disadvantage of this feature is the fact that its computation may imply a division by a small number which is susceptible to statistical variations or variations caused by artifacts. Furthermore, the appropriateness of this normalisation for lesion detection has been challenged in literature. Wismüller et al., 2002 point out that the initial identification of lesions primarily depends on the high amplitude of CA uptake, whereas the dynamics of the signal predominantly provide information about subclasses of lesion tissue. Nevertheless, this signal normalisation was already employed for evaluating DCE-MRI signals using a multi-layer perceptron [Lucht et al., 2002, Lucht et al., 2001].

- The *prenorm*-feature

$$\mathbf{x_p} = (s_{\mathbf{p}2} - s_{\mathbf{p}1}, \ldots, s_{\mathbf{p}n_t} - s_{\mathbf{p}1}) \in \mathcal{X} = \mathbb{R}^{n_t-1} \tag{5.6}$$

describing the amount of signal enhancement between the signal value in the precontrast image and in one of the postcontrast images. Also the amount of signal enhancement as a basis for a quantitative signal assessment, in particular for the purpose of distinguishing benign and malignant signals, has been questioned due to its vulnerability to local variations of the coil reception and inter-subject variations [Heywang-Köbrunner and Beck, 1995]. On the other hand, these features provide the same information as visualised by the subtraction images computed from the precontrast and the $n_t - 1$ postcontrast images.

- The *allratio*-feature

$$\mathbf{x_p} = \left(\frac{s_j}{s_k}\right) \in \mathcal{X} = \mathbb{R}^{\frac{(n_t-1)n_t}{2}} \tag{5.7}$$

with $j, k = 1, \ldots, n_t$ and $k < j$. This feature reflects the ratios of intensity values at two different points in time j and k. Therewith, the *allratio*-feature can be regarded as an extension of the *preratio*-feature which may lead to a description of the signal course which is more suitable for evaluation by an ANN.

Region-Based Temporal Texture Features

Textural features descriptors are widely used in the field of computer vision and medical image analysis [Castellano et al., 2004]. The texture of a region or window is a periodically repeating spatial pattern of gray value variations. By means of textural features, larger regions showing certain periodical patterns can be considered as homogenous. Textural features have been used for many medical problems such as texture-based image segmentation or lesion detection in image data obtained from modalities such as X-ray, ultrasound or MRI [Bankman, 2000]. In the domain of MR imaging, textural features have been used for instance for tissue classification in MR images of brains [Herlidou-Meme et al., 2003, Kmer et al., 1995], classification of lesions in high-resolution DCE-MRI images of breasts [Gibbs and Turnball, 2003] or for detecting simulated microcalcifications in DCE-MRI images [James et al., 2001]. The basic assumption behind textural approaches for tissue classification is that the varying structure of different tissue types is reflected in its visual appearance and exhibits tissue-specific texture patterns. If this assumption holds, the texture of e.g. rectangular windows can be assessed using dedicated mathematical techniques. The resulting vector of textural features is then assigned to the pixel at the centre of the window and provides a description of the pixel's neighbourhood. By moving the window over the entire image, a feature vector is calculated for each pixel. Thereby, windows displaying the same tissue are likely to yield similar feature vectors.

Due to the central role of textures in computer vision, a wide range of methods for the mathematical description of features has been proposed such as *wavelet* [Daubechies, 1991, Aldroubi and Unser, 1996] and *Gabor* filters (e.g. [Grigorescu et al., 2002, Fogel and Sagi, 1989]), model-based techniques using e.g. *Markov-random-fields* or statistical approaches describing the first and second order statistics of gray values (e.g. [Haralick et al., 1973]). One of the most frequently used textural description, which has also been applied for the analysis of MRI data [James et al., 2001], are Haralick's features derived from the co-occurrence matrix describing second order statistics of gray level values [Haralick et al., 1973]. Originally introduced to describe two-dimensional images, they have recently been extended for processing temporal sequences of three-dimensional images [Woods et al., 2004]. Nevertheless, it is not uncommon to handle three-dimensional MR images as a set of two-dimensional images, particular if voxels expose an anisotropic shape, i.e. an asymmetrical extent along the three spatial axis.

For a reliable assessment of the texture of a region, the windows for which texture features are computed have to be of a reasonable size. They have to be sufficiently large to display the periodicity of the pattern. On the other hand, large windows are more likely to show more than one texture. In this case, the computed feature vector relates to more than one texture type and tends to be less discriminative. In particular for detecting arbitrary shaped tissue structures such as lesion masses, the size of the rectangular window should not significantly exceed the extent of lesions.

For the task of lesion detection as presented in this chapter, the textural feature description $\mathbf{x_p}$ relates to a 5×5 two-dimensional image window $\mathbf{W_p}$ with centre pixel $\mathbf{p}$. The three-dimensional image data is processed as a sequence of two-dimensional images whose orientation are chosen in such manner that displayed pixels expose a symmetrical shape (*Munich* group: Axial orientation, *MARIBS* group: Coronal orientation). Textural feature descriptors are calculated using *Law's filter masks* describing the texture energy [Laws, 1979]. Based on the five one-dimensional convolution

masks *level*, *edge*, *spot*, *wave* and *ripple*

$$\begin{aligned}
\mathsf{L5} &= (+1,+4,+6,+4,+1)\ \ (\mathsf{Level}),\\
\mathsf{E5} &= (-1,-2,+0,+2,+1)\ \ (\mathsf{Edge}),\\
\mathsf{S5} &= (-1,+0,+2,+0,-1)\ \ (\mathsf{Spot}),\\
\mathsf{W5} &= (-1,+2,+0,-2,+1)\ \ (\mathsf{Wave}),\\
\mathsf{R5} &= (+1,-4,+6,-4,+1)\ \ (\mathsf{Ripple}),
\end{aligned}$$

25 two-dimensional convolution masks are determined by calculating the outer-product of all permutations of the one-dimensional masks $\mathbf{b}_1, \mathbf{b}_2 \in \{\mathsf{L5}, \mathsf{E5}, \mathsf{S5}, \mathsf{W5}, \mathsf{R5}\}$. 14 directional-invariant *Law's texture energy measure*s for the window $\mathbf{W}$ are calculated by

$$v^{\mathbf{b}_1\mathbf{b}_2} = \frac{g^{\mathbf{b}_1\mathbf{b}_2} + g^{\mathbf{b}_2\mathbf{b}_1}}{g^{L5L5}},$$

where

$$g^{\mathbf{b}_1\mathbf{b}_2} = \sum_{m}^{5}\sum_{n}^{5}(|(\mathbf{b}_1\mathbf{b}_2^T)\cdot\mathbf{W}|)_{m,n},\ \ \mathbf{b}_1, \mathbf{b}_2 \in \{\mathsf{L5}, \mathsf{E5}, \mathsf{S5}, \mathsf{W5}, \mathsf{R5}\}.$$

The fifteenth feature v^{L5L5}, which reflects the centre-weighted local average, is typically only used for normalisation of the other feature components and is subsequently discarded. In order to consider the temporal dimension of the DCE-MRI images, the textural features are calculated for two subtraction images. The first subtraction image is calculated by subtracting the precontrast image from the first postcontrast image. The second subtraction image is calculated using the last postcontrast image. Thus, the textural information about the content of the window $\mathbf{W}$ is represented by a vector $\mathbf{v} \in \mathbb{R}^{n_{\mathsf{Laws}}}$ of $n_{\mathsf{Laws}} = 28$ features.

To avoid the phenomenon known as *curse of dimensionality*, which was already mentioned in chapter 4, the dimensionality of the textural feature vectors is reduced by *principal component analysis* (PCA) [Bishop, 1995] of the training data. First, each textural feature is normalised to zero mean and unit variance. Subsequently, the eigenvectors $\boldsymbol{\xi}_l,\ l = 1, \ldots, n_{\mathsf{Laws}}$ and the corresponding eigenvalues λ_l of the covariance matrix

$$\mathbf{C} = \sum_{i=1}^{N} \mathbf{v}_i \mathbf{v}_i^T$$

are calculated. By sorting the eigenvectors in the order of descending eigenvalues, one can create an ordered orthogonal basis with the first eigenvector reflecting the direction of largest variance of the data. The ratio $\lambda_l / \sum_{k=1}^{n_{\mathsf{Laws}}} \lambda_k$ specifies the fraction of data variance retained by projecting onto the l-th eigenvector. Assuming that the variance of the smaller eigenvalues mainly represents noise, the eigenvectors corresponding to the n_{in} largest eigenvalues are selected retaining approximately 90% of the variance of the textural data. The n_{in} eigenvectors are combined to the transformation

matrix $\mathbf{A} \in \mathbb{R}^{n_{\text{in}} \times n_{\text{Laws}}}$ defining a linear mapping from the textural feature space $\mathbb{R}^{n_{\text{Laws}}}$ to the data space $\mathcal{X}$, which has $n_{\text{in}} = 18$ dimensions in the case of the *MARIBS* data and $n_{\text{in}} = 19$ dimensions in the case of the *Munich* data. The final feature vectors $\mathbf{x_p}$, which are subsequently evaluated by the supervised classification algorithm, are computed by

$$\mathbf{x_p} = \mathbf{A}\mathbf{v_p}.$$

The reader is referred to chapter 8 for more details about PCA.

5.1.4 Adaptation of Predictors

Support Vector Machine

The SVM algorithm is applied in combination with two different kernel functions. In combination with a linear kernel, adaptation of the SVM leads to a discrimination function dividing the data space $\mathcal{X}$ by a hyperplane into two half-spaces. Since the linear kernel has no additional kernel-specific hyperparameters, the only hyperparameter that needs to be selected by the user is the regularisation parameter C controlling the trade-off between maximisation of the margin and minimisation of the training error. In the case of the Gaussian kernel, the examples of the two classes are distinguished by a non-linear discrimination function. Next to the regularisation parameter C, the kernel bandwidth σ has to be selected. The values of these hyperparameters are chosen by a grid search in the one- or two-dimensional hyperparameter space, respectively (see chapter 4.2.3). To this end, the data set $\Gamma = \Gamma^{\text{Suspicious}} \cup \Gamma^{\text{Normal}}$ sampled from the training cases is divided into two disjoint subsets

$$\Gamma_{\text{Train}} = \Gamma^{\text{Suspicious}}_{\text{Train}} \cup \Gamma^{\text{Normal}}_{\text{Train}}$$

and

$$\Gamma_{\text{Test}} = \Gamma^{\text{Suspicious}}_{\text{Test}} \cup \Gamma^{\text{Normal}}_{\text{Test}}.$$

The set of suspicious training examples $\Gamma^{\text{Suspicious}}_{\text{Train}}$ contains 50%, but not more than 4000 examples, of $\Gamma^{\text{Suspicious}}$. The remaining examples are used to build the test set $\Gamma^{\text{Suspicious}}_{\text{Test}} = \Gamma^{\text{Suspicious}} / \Gamma^{\text{Suspicious}}_{\text{Train}}$. These sets are then combined with normal examples randomly selected from Γ^{Normal}. The number of normal examples is chosen such that the final training set Γ_{Train} is balanced, i.e. it contains an equal number of normal and suspicious examples, whereas Γ_{Test} contains ten times as much normal as suspicious examples, but no more than 20000 suspicious examples. The size of Γ_{Test} and Γ_{Train} is bounded in order to limit the computational expenses of the grid-search. Each feature of the training examples is normalised to zero mean and unit variance by a z-scoring operation. This operation is also applied to Γ_{Test}, respectively.

During grid-search, the performance of the SVM for the hyperparameter values as shown in table 5.1 is assessed by ROC analysis (see chapter 4.5.2). The score values of the examples in Γ_{Test}, i.e. the distances of the examples to the hyperplane in the kernel-induced feature space $\mathcal{F}$,

are utilised to calculate a ROC curve which is subsequently summarised by the *area-under-the-ROC-curve* (Az) value. The DCE-MRI sequence of the unseen case is then evaluated using the SVM trained with the hyperparameters yielding the highest Az value on the test set.

Visualisation of SVM Outcome

To facilitate visualisation of the resulting confidence volume, the output $f_{\boldsymbol{\theta}}(\mathbf{x})$ of the trained SVM is transformed by a sigmoid-function to the a-posteriori probability $P(\text{Suspicious}|\mathbf{x}) \in [0;1]$ using the method of [Platt, 2000] (see chapter 4.2.5). Through this, the outcome of the classifier can be displayed as a gray value image by multiplying the a-posteriori probability $P(\text{Suspicious}|\mathbf{x})$ computed for the voxels of a DCE-MRI sequence with a constant factor of 255. The parameters of the sigmoid function, which are fitted in a calibration process to the scores $f_\theta(\mathbf{x}) \in \Gamma_{\text{Train}}$ computed by five-fold cross validation, are estimated using the algorithm proposed by Lin et al., 2003. Since the sigmoid function is fitted to a balanced data set, the returned a-posteriori probability is based on the assumption of equal a-priori probabilities for both tissue classes. Even though the true a-priori probabilities are unknown, one can assume that the a-priori probability for suspicious signals is lower than for normal signals. In order to consider unequal priors and to use similar a-priori probabilities as for the LDA based evaluation of the image data, the a-priori probabilities $P_t(\text{Normal})$ and $P_t(\text{Suspicious})$ (t indicates estimates based on the training set) are adjusted to $P(\text{Normal}) = 0.9$ and $P(\text{Suspicious}) = 0.1$ using the *Bayes formula*:

$$p(\mathbf{x}|\omega_k) = \frac{P(\omega_k|\mathbf{x})P(\mathbf{x})}{P(\omega_k)}. \tag{5.8}$$

Since the within-class densities $p_t(\mathbf{x}|\omega_k)$ estimated from the training set do not change during the correction step, one can equate the estimated density and the corrected density:

$$p_t(\mathbf{x}|\omega_k) = p(\mathbf{x}|\omega_k). \tag{5.9}$$

Substitution of the Bayes formula (5.8) leads to

$$\frac{P_t(\omega_k|\mathbf{x})P_t(\mathbf{x})}{P_t(\omega_k)} = \frac{P(\omega_k|\mathbf{x})P(\mathbf{x})}{P(\omega_k)},$$

which finally leads to the formula for the corrected a-posteriori probability of class ω_k:

$$P(\omega_k|\mathbf{x}) = \frac{\frac{P(\omega_k)}{P_t(\omega_k)}P_t(\omega_k|\mathbf{x})}{\sum_{j=1}^{n_\Omega}\frac{P(\omega_j)}{P_t(\omega_j)}P_t(\omega_j|\mathbf{x})}$$

Note that the postprocessing of the SVM output by the monotonic sigmoid function as well as the correction of the assumed a-priori probabilities does not change the ranking of the evaluated examples ordered according to the postprocessed confidence value. Thus, both postprocessing steps do not influence ROC analysis which only considers the ranking of examples. The main

Table 5.1: Range of values for the regularisation parameter C and kernel parameter σ which is scanned for selecting the optimal SVM hyperparameters.

Kernel Type	Regularisation	Kernel
Linear Kernel	$C = 10^{i0.2}, i = 1, \ldots, 10$	none
Gaussian Kernel	$C = 10^{i0.2}, i = 1, \ldots, 10$	$\sigma = 0.5, \ldots, 5$, 10 steps

purpose of postprocessing is to enhance the initial visual contrast between voxels which are rated as suspicious with moderate to high confidence. Applying the sigmoid function maps the unbounded range of score value to a fixed range of confidence values. These confidence values can then directly be mapped to gray values. The correction of the a-priori probabilities improves the mapping for voxels with moderate and high confidence values. After correction, low confidence values are mapped to a smaller range of gray values, whereas moderate to high confidence values are displayed by an increased number of nuances of gray values. Even though the same effect can be obtained by manually tuning the mapping as it is routinely performed by most radiologists, the outcome depends on the radiologist's expertise and is a potential source of inter- and intra-observer variance. Therefore, already the initial display of the confidence volume should enable the observer to distinguish lesion tissue from normal tissue as good as possible.

Linear Discriminant Analysis

LDA, like the SVM-L, divides the data space by a hyperplane into two half-spaces. In contrast to the SVM, no hyperparameter has to be selected which makes the training process for the LDA computationally inexpensive. The linear discrimination function is determined by assuming normally distributed classes with equal covariances. The parameters, i.e. the class-specific means $\boldsymbol{\mu}_{\text{Normal}}$, $\boldsymbol{\mu}_{\text{Suspicious}}$, the common covariance matrix Σ of the pooled data for the class-conditional densities $p(\mathbf{x_p}|\text{Suspicious})$ and $p(\mathbf{x_p}|\text{Normal})$ and the a-priori probabilities $P(\text{Normal})$ and $P(\text{Suspicious})$, are estimated on a training set Γ_{Train}. This set consists of all suspicious signals of training cases and ten times as much randomly selected normal signals. Evaluation of an unseen DCE-MRI assigns the a-posteriori probability

$$P(\text{Suspicious}|\mathbf{x_p}) = \frac{p(\mathbf{x_p}|\text{Suspicious})P(\text{Suspicious})}{p(\mathbf{x_p}|\text{Suspicious})P(\text{Suspicious}) + p(\mathbf{x_p}|\text{Normal})P(\text{Normal})}$$

to the corresponding voxels in the confidence volume, which therefore can directly be visualised as a three-dimensional gray value image.

5.2 Results

In order to assess the performance of the different combinations of classification algorithms and feature types, the cases of the *Munich* group and the cases of the *MARIBS* group are evaluated following a *leave-one-case-out* scheme. For a pool of M cases, each setup, i.e. a combination of classifier and feature type, is trained with data sampled from $M - 1$ cases and tested on the remaining unseen case. This process is repeated five times for all permutations of test and training

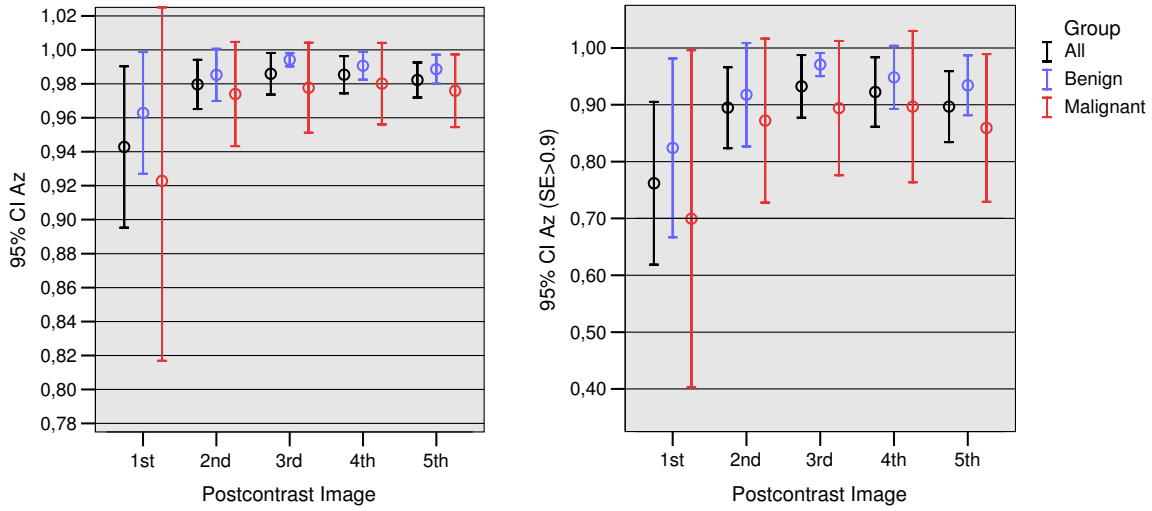


Figure 5.5: *Area-under-ROC curve* (Az) values (left) and *partial-area-under-ROC-curve above a sensitivity of* 0.9 ($\text{Az}_{(\text{SE}>0.9)}$) values (right) for subtraction images based on the postcontrast images as measured at five different points of time. Each plot represents the means and the corresponding 95% confidence intervals calculated for the subgroup of benign cases (blue), the subgroup of malignant cases (red) and the entire *Munich* group.

cases leading to $M \times 5$ confidence volumes for each combination of classifier and feature type. For each confidence volume, the *area-under-ROC-curve* (Az) and *partial-area-under-ROC-curve above a sensitivity of 0.9* ($\text{Az}_{(\text{SE}>0.9)}$) are calculated (see chapter 4.5.2). The performance of the different setups is then compared by means of the average Az and $\text{Az}_{(\text{SE}>0.9)}$ values.

5.2.1 Results for the *Munich* Group

The *Munich* group consists of six cases with histologically proven benign lesions and six cases with histologically proven malignant lesions. The lesion size varies from 25 to 743 voxels with an average size of 208. Details about the image acquisition and the outcome of the histological examination are described in chapter 3.

Subtraction Images

Figure 5.5 visualises the means and the corresponding 95% confidence intervals for the Az and $\text{Az}_{(\text{SE}>0.9)}$ values of the subtraction images based on the precontrast image and on one of the five postcontrast images of the *Munich* cases. Considering all twelve cases simultaneously, the subtraction images based on the third postcontrast image yields the best performance ($\text{Az} = 0.986 \pm 0.0192$, $\text{Az}_{(\text{SE}>0.9)} = 0.932 \pm 0.0868$). The same subtraction image yields also the best results for the subgroup of benign cases ($\text{Az} = 0.994 \pm 0.0037$, $\text{Az}_{(\text{SE}>0.9)} = 0.971 \pm 0.0195$). In contrast to the initial assumption that a subtraction image based on one of the early postcontrast images is most suitable for the detection of the enhancing malignant tissue, the highest mean value is yielded for the subtraction image based on the fourth postcontrast image ($\text{Az} = 0.980 \pm 0.0228$,

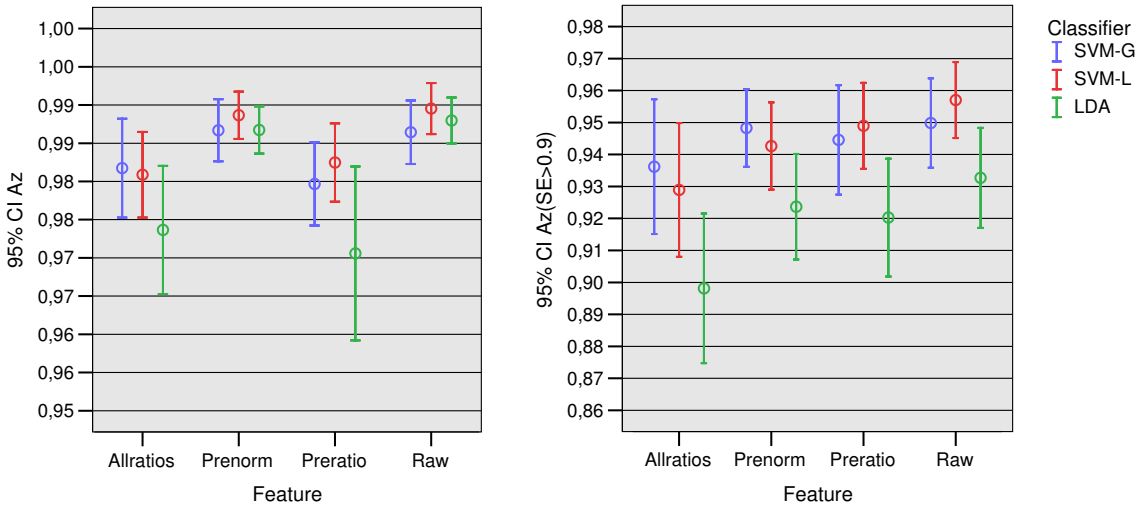


Figure 5.6: Area-under-ROC curve (Az) values (left) and partial area-under-ROC-curve above a sensitivity of 0.9 ($\mathrm{Az}_{(\mathrm{SE}>0.9)}$) values (right) for different voxel-based temporal kinetic features combined with the SVM-G (blue), SVM-L (red) and LDA (green). The mean values and the corresponding 95% confidence intervals are calculated by averaging the 60 values obtained from the 5 repetitions of the leave-one-case-out setup.

$\mathrm{Az}_{(\mathrm{SE}>0.9)} = 0.897 \pm 0.1268$). While the values of both indices are only slightly different to those of the subtraction images based on the second or third postcontrast images, they are superior to the mean value of the subtraction image based on the first postcontrast image ($\mathrm{Az} = 0.923 \pm 0.1001$, $\mathrm{Az}_{(\mathrm{SE}>0.9)} = 0.6995 \pm 0.2826$). In summary, the similar mean values measured for the subtraction images based on the later postcontrast images indicate that all these images provide information about the localisation and extent of suspicious masses. These images should therefore be considered by the radiologist for an accurate and comprehensive evaluation of the DCE-MRI data.

Voxel-Based Temporal Kinetic Features

Figure 5.6 represents the means and 95% confidence intervals of Az (left) and $\mathrm{Az}_{(\mathrm{SE}>0.9)}$ (right) for different combinations of voxel-based temporal kinetic features and classification algorithms. For all four types of features, the SVM-L outperforms the LDA. Employing a Gaussian kernel, which extends the class of discrimination functions to non-linear functions, does not lead to significant improvements of the detection performance. An Az of 0.989 ± 0.0129 and $\mathrm{Az}_{(\mathrm{SE}>0.9)}$ of 0.957 ± 0.0459 indicates that the SVM-L evaluating the unprocessed temporal kinetic signal (*raw*-feature) is most suitable for the assessment of temporal kinetic signals. In particular, the $\mathrm{Az}_{(\mathrm{SE}>0.9)}$ indicates an increased sensitivity in comparison with the best subtraction image.

Axial slices of the subtraction images based on the precontrast and third postcontrast image for two benign cases B1 and B4 and two malignant cases M4 and M6 can be observed in the left image column of figure 5.7. The intensity values of voxels marked by the breast mask,

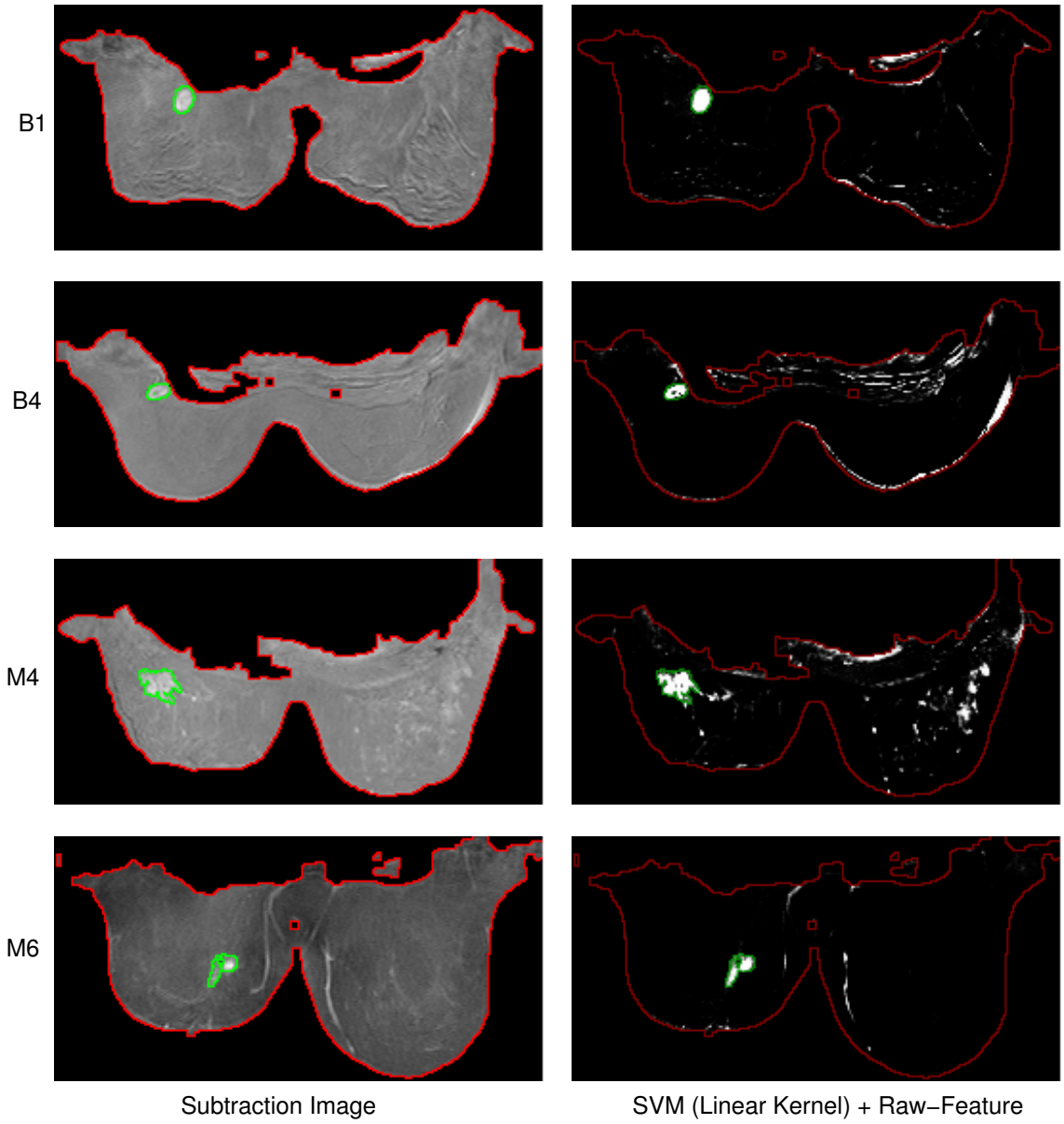


Figure 5.7: **Left column:** Axial slices of the subtraction images of two benign cases B1 and B4 and two malignant cases M4 and M6 of the *Munich* group. Each image visualises an axial slice of the three-dimensional subtraction image based on the third postcontrast image, whose voxel values are scaled to the range $[0; 255]$ for the purpose of visualisation. Only voxels marked by the breast mask, whose contour is indicated by the red line. The green line depicts the contour of the lesion segmentation. **Right column:** The same axial slices as visualised in the left column, but with voxel intensities reflecting the confidence values returned by the SVM-L evaluating the *raw*-feature description.

whose contour is indicated by the red line, are scaled to the range of $[0; 255]$ for the purpose of visualisation. In each image, a green line marks the contour of the lesion as segmented by

the radiologist. All four lesions appear as clusters of voxels with intensity values higher than those of voxels of the surrounding tissue. The bifocal lesion of case M6 consists of two regions. The left region predominantly exhibits temporal kinetic signals with benign signal characteristics, whereas the right region is likely to be malignant according to the temporal kinetic signals. In the subtraction image, voxels of the malignant subregion can clearly be distinguished from voxels of the surrounding normal tissue. The benign region exhibits only a slight signal enhancement and is therefore displayed by voxels with lower intensities. Next to the lesion, major blood vessels are also displayed by bright voxels. The subtraction image for case M4 also exposes several diffusely enhancing tissue regions in the left breast, whereas the lesion, according to the expert label, is located in the right breast.

The right column of figure 5.7 depicts the same slices as the left column, but voxel intensities reflect the confidence value returned by the SVM-L evaluating the unprocessed temporal kinetic signal (*raw*-feature). For the purpose of visualisation, each confidence value is multiplied by a constant value of 255. All four lesions are displayed as clusters of bright voxels. The extents of the clusters match the contours of the corresponding expert labels. In contrast to the subtraction image, both parts of the bifocal lesion of case M6 can clearly be distinguished from normal tissue, which is displayed in dark colours in all four images. Beside designated lesions, additional regions are spuriously highlighted as suspicious in all four images. In the confidence image computed for case B1, spuriously highlighted voxels in the breast can roughly be grouped in horizontal lines. Since these lines are perpendicular to the direction of respiration and cardiac motions, they are likely to be caused by the corresponding motion artifacts in the image data. The extents of the tissue regions in the left breast of case M4, which are displayed as diffusely enhancing regions in the subtraction image, are more clearly defined in the SVM-L based visualisation. Even though the regions were not marked by the radiologist, examination of the corresponding temporal kinetic signals shows that the underlying tissue exhibits a considerable signal enhancement within the first two postcontrast images. The characteristics of the temporal kinetic signals of these spuriously highlighted subregions will be investigated in more detail in the following chapter. Since the temporal courses of the corresponding signals exhibit characteristics of benign and malignant tissue as well, the highlighting of the voxels by the classifier seems to be reasonable from the point of view of an assessment which solely considers the temporal kinetic signals associated with individual voxels.

Analysis of the Discrimination Function

For the SVM-L evaluating the *raw*-feature, additional knowledge about the learned model of tissue signals can be derived from the weight vector

$$\mathbf{w} = \sum_{i=1}^{N} y_i \alpha_i \mathbf{x}_i$$

describing the parametrisation of the discrimination function. Since each feature component of the training data is normalised to zero mean and unit variance, the absolute value of the corresponding component of the weight vector describes the impact of the feature component on the outcome of the discrimination function. For the *raw*-feature, components of the feature vector

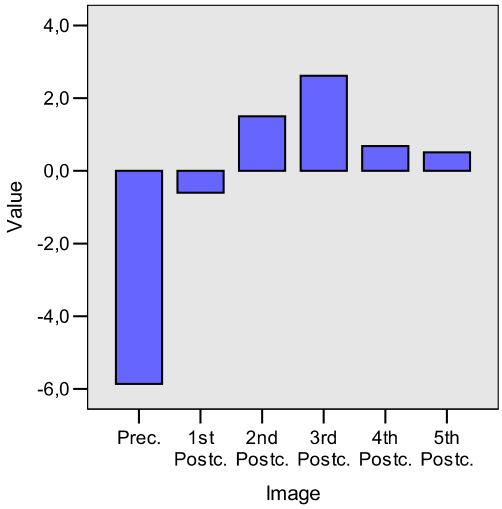


Figure 5.8: Weight vector $\mathbf{w}$ determined by the SVM-L for the evaluation of the *raw*-feature. The absolute value of the component $w_k, k = 1, \ldots, n_{in}$ of the weight vector reflects the impact of the feature x_k on the classification outcome.

Table 5.2: Mean values and standard deviations for the ROC measures Az and $\text{Az}_{(\text{SE}>0.9)}$ for the evaluation of the *Munich* group images based on region-based textural features.

	Az	**Az**$_{(\textbf{SE}>0.8)}$
SVM-G	0.8993 ± 0.0830	0.6636 ± 0.2159
SVM-L	0.9060 ± 0.0779	0.6802 ± 0.2246
LDA	0.9113 ± 0.0741	0.6901 ± 0.2190

correspond to the intensity of a voxel as measured in the pre- and postcontrast images. Therefore, the corresponding components of the weight vectors illustrate the value of the different images for the discrimination of normal and suspicious tissue. According to the weight vector displayed in figure 5.8, the outcome of the classification function predominantly depends on the intensity value in the precontrast and in the third postcontrast image. This finding is congruent with the observation that the highest discriminative power is achieved by evaluating the subtraction image displaying difference between the precontrast and the third postcontrast image.

Region-Based Temporal Texture Features

Table 5.2 depicts the Az and $\text{Az}_{(\text{SE}>0.9)}$ values and the corresponding standard deviations yielded by the three classifiers evaluating the Law's features computed for the 5×5 neighbourhood of each voxel. The best results $\text{Az} = 0.9113 \pm 0.0741$ and $\text{Az}_{(\text{SE}>0.9)} = 0.6901 \pm 0.2190$ are obtained using the LDA, but, when taken into account the high standard deviation, both indices are similar to those of the SVM-L and SVM-G. In comparison with the voxel-based temporal kinetic features ($\text{Az} = 0.989 \pm 0.0129$, $\text{Az}_{(\text{SE}>0.9)} = 0.957 \pm 0.0459$), the ROC analysis clearly indicates an inferior performance for the setup based on the evaluation of textural features.

5.2.2 Results for the *MARIBS* Pool

The *MARIBS* group consists of three cases with benign lesions and five cases with malignant lesions. The DCE-MRI sequences differ from those of the *Munich* group in two aspects:

1. The lesions of the *MARIBS* cases are significantly larger than those of the *Munich* pool. The size of the lesions ranges from 175 voxels to 4307 voxels. In total, the eight cases contain 12022 voxels designated by the radiologist as lesion voxels. Hence, the application of the SVM algorithm not only requires to under-sample the class of normal signals but also the class of suspicious signals. Due to the leave-one-case-out evaluation setup, which requires to repeat the grid search for optimal hyperparameters for each validation case, the SVM algorithm has to be executed 130 times for each combination of validation case and feature type. Even if sophisticated algorithms are applied for SVM training, the number of suspicious examples considered during the training of the SVM has to be limited to 4000, which is about 25% of the available number of suspicious examples. In contrast to the SVM, the LDA can be adapted using the entire set of suspicious examples.

2. According to the diagnosis-form-sheet, several malignant cases exhibit a large number of secondary regions of benign disorders. Since the radiologist only segmented the primary disorder, such secondary regions have to be considered as normal tissue during training and evaluation and may cause a considerable overlap between the data distributions of both tissue classes.

For the following ROC analysis of the subtraction images and confidence volumes, the case B013A will be excluded. The Az values of the subtraction images of case B013A are ranging between 0.3776 to 0.5924 and, therewith, are exceptional low in contrast to the indices of the remaining cases. Therefore, B013A is regarded as an outlier and will be discussed separately from the group of the remaining cases. Note that B013A is nonetheless part of the leave-one-case-out scheme and, therewith, provides data for adapting the classifiers.

Subtraction Images

Figure 5.9 illustrates the means Az and $\mathrm{Az}_{(\mathrm{SE}>0.9)}$ and the corresponding 95% confidence intervals for the subtraction images based on the postcontrast images measured at five different points of time. According to the mean Az value, the subtraction image based on the last postcontrast image yields the highest Az value of 0.9389 ± 0.0250 and $\mathrm{Az}_{(\mathrm{SE}>0.9)} = 0.6002 \pm 0.1604$. Nevertheless, these results are comparable to those of the subtraction images based on the third postcontrast image ($\mathrm{Az} = 9.9346 \pm 0.0267$ and $\mathrm{Az}_{(\mathrm{SE}>0.9)} = 0.5819 \pm 0.1337$) or on the fourth postcontrast image ($\mathrm{Az} = 0.9367 \pm 0.0274$ and $\mathrm{Az}_{(\mathrm{SE}>0.9)} = 0.5929 \pm 0.17.01$).

Voxel-Based Kinetic Features

Figure 5.10 illustrates the mean Az and $\mathrm{Az}_{(\mathrm{SE}>0.9)}$ and the corresponding 95% confidence intervals as yielded by the different combinations of classification algorithms and voxel-based temporal kinetic features. For the *preratio*-feature and *allratio*-feature, the SVM-G outperforms both linear classification algorithms. The best performance is achieved by employing the SVM-G for the

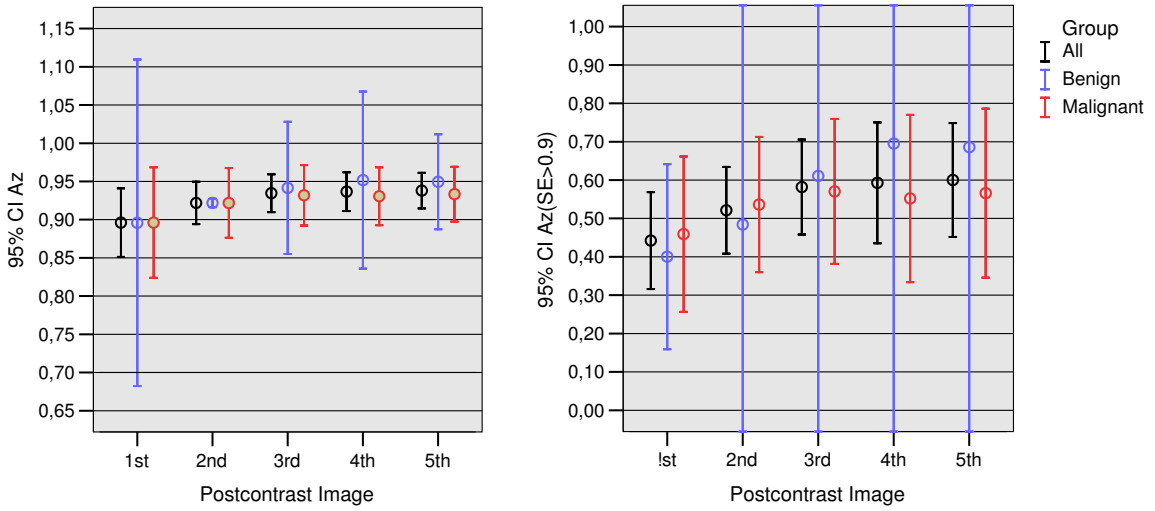


Figure 5.9: Mean values and corresponding confidence intervals for the *area-under-ROC-curve* (Az) (left) and *partial-area-under-ROC-curve above a sensitivity of 0.9* ($Az_{(SE>0.9)}$) (right) for the subtraction images of the cases of the *MARIBS* group excluding case B013A. The mean values are separately calculated for the subtraction images based on the postcontrast images measured at five different points of time. Further on, the group-specific mean values for the group of malignant cases (red), of benign cases (blue) and of all cases (black) are determined.

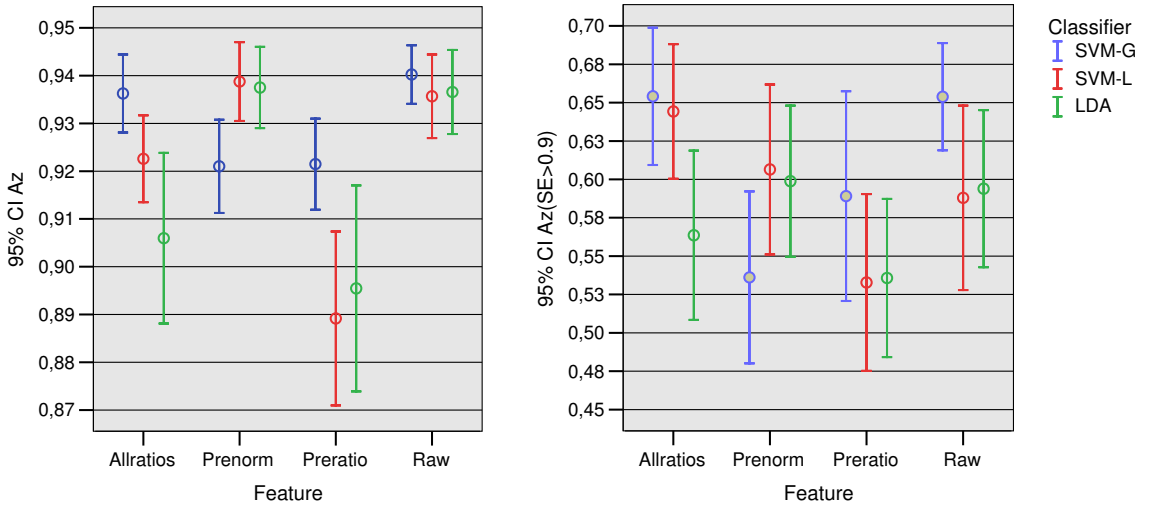


Figure 5.10: Mean values with 95% confidence intervals of the ROC measures Az (left) and $Az_{(SE>0.9)}$ (right) for different combinations of classification algorithms and voxel-based temporal kinetic features applied for detecting suspicious signals in image volumes of the *MARIBS* group.

evaluation of the unprocessed temporal kinetic signal (*raw*-feature) leading to an $Az = 0.94203 \pm 0.0178$ and $Az_{(SE>0.9)} = 0.6539 \pm 0.1017$. Therewith, the machine learning based evaluation of

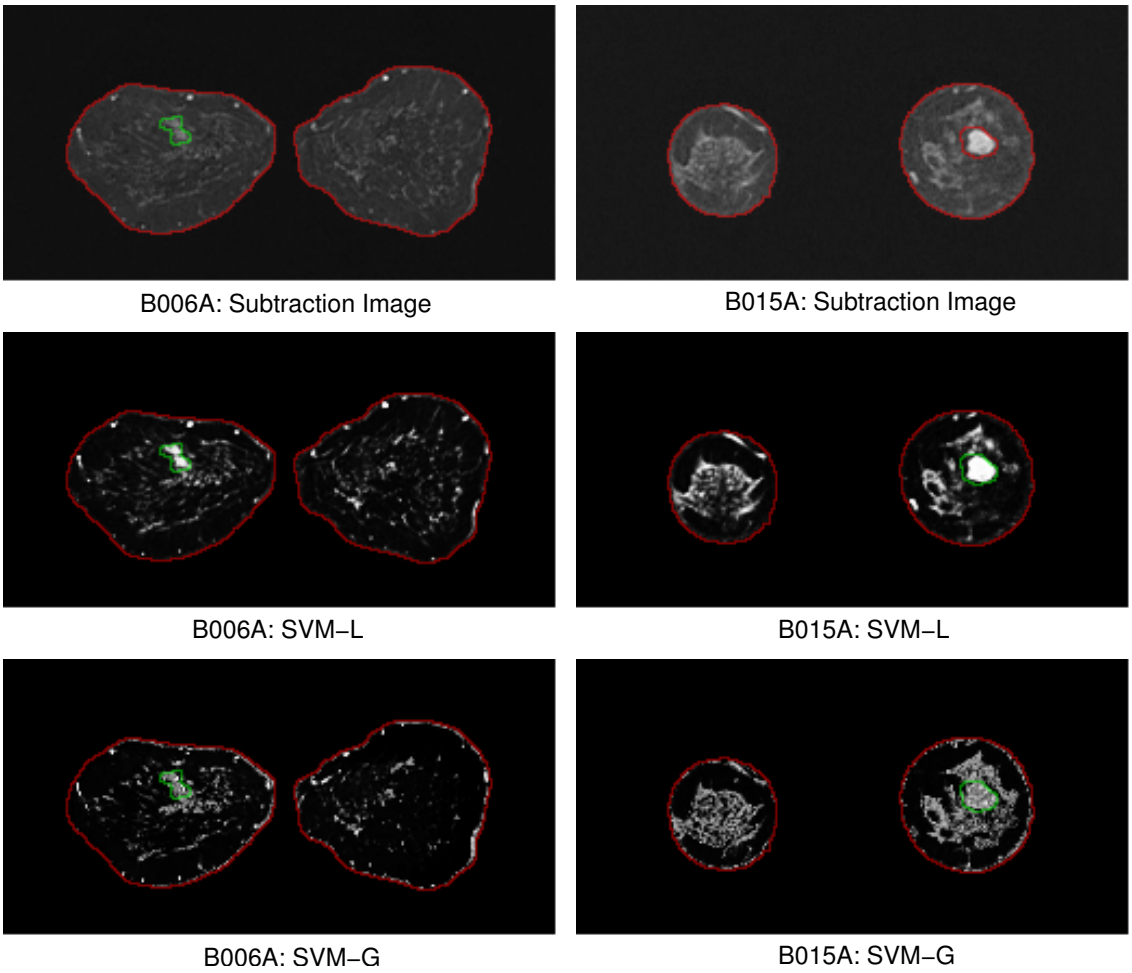


Figure 5.11: Subtraction images and different confidence volumes for the two benign cases B006A (left column, disorders; fibrocystic changes, sclerosing lymphocytic lobulitis) and B015A (right column, disorders: fibroadenoma, fibrocystic changes, apocrine metaplasia) of the *MARIBS* group. Each column depicts the same coronal slice extracted from the subtraction image based on the third postcontrast image, from the confidence volume based on the SVM-L and from the confidence volume based on the SVM-G (lower left). The latter two were computed by evaluating the unprocessed temporal kinetic signal (*raw*-feature). Voxel values of the subtraction images are scaled to the range $[0; 255]$. Confidence values returned by the SVM are mapped to gray values by multiplication with 255. The contour of the expert label of the lesion is depicted by the green line. The red line illustrates the contour of the breast mask.

the DCE-MRI data leads to an improvement of the detection performance compared to the best subtraction image.

Figure 5.11 shows coronal slices of the subtraction images and confidence volumes for the two benign cases B006A (left column of images) and B015A (right column of images). The confidence volumes were computed by evaluating the *raw*-feature using the SVM-G or the SVM-L. The contours of the lesion and the breast mask are reflected by a green and red line, respectively.

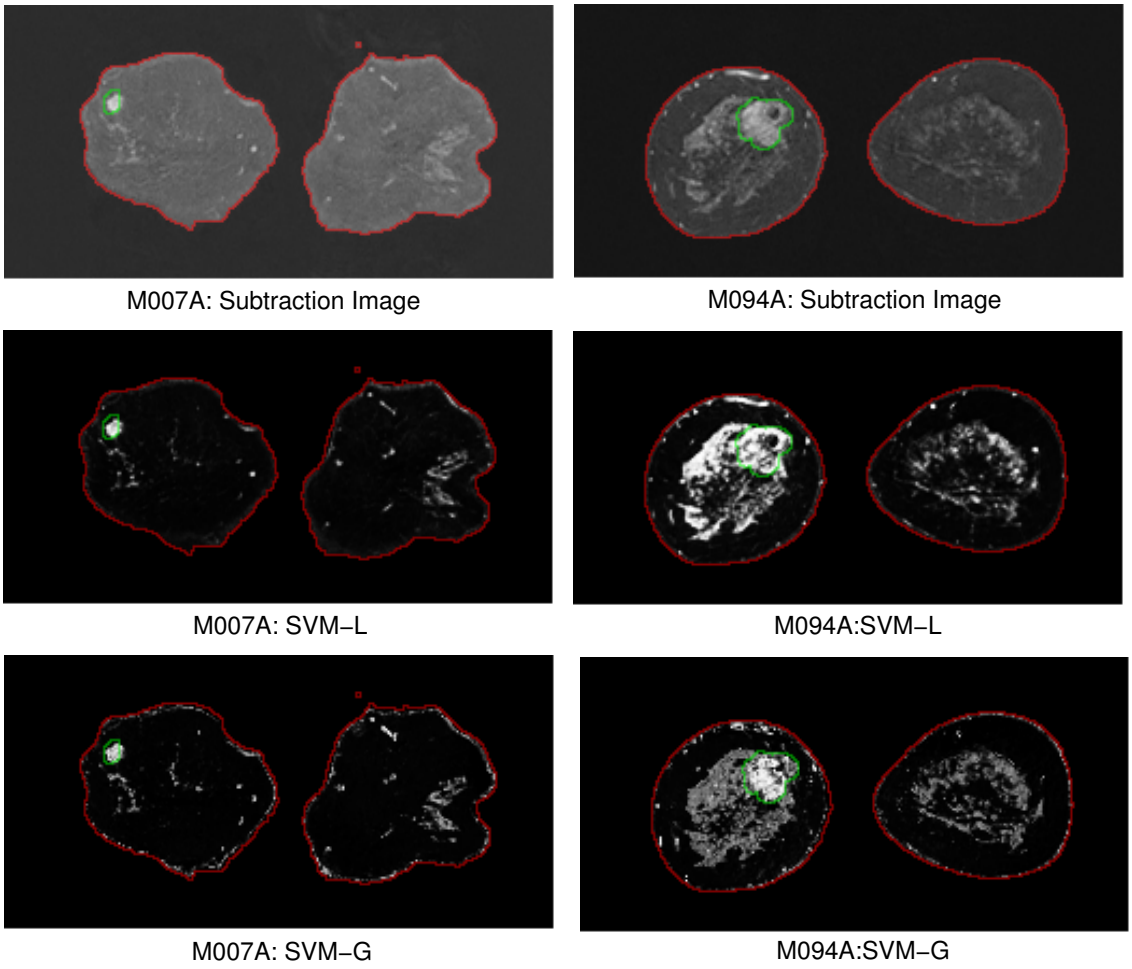


Figure 5.12: Subtraction images and different confidence volumes for the two malignant cases M007A (Left column, primary disorder: Non-invasive ductal carcinoma, low grade; invasive tubular carcinoma grade I; secondary disorders: Fibrocystic change, periductal mastitis, sclerosing adenosis, other) and M094A (Right column, primary disorder: Non-invasive ductal, invasive ductal, grade III) of the *MARIBS* pool. Each column depicts the same slice extracted from the subtraction image based on the third postcontrast image, from the confidence volume based on the SVM-L and from the confidence volume based on the SVM-G. All confidence volumes are based on the evaluation of the unprocessed temporal kinetic signal (*raw*-feature). Voxel values of the subtraction images are scaled to the range $[0; 255]$. Confidence values returned by the SVM are mapped to gray values by multiplication with 255. The contour of the expert label of the lesion is depicted by the green line. The red line illustrates the contour of the breast mask.

The subtraction image of case B006A displays the lesion as a weakly enhancing region. Slightly enhancing regions are also spread over the entire breast. The strongest signal enhancement can be observed for small circular regions near the skin and is caused by larger blood vessels perpendicularly oriented to the image slice. The SVM-L based confidence image displays the lesion as a cluster of bright voxels. Spuriously highlighted regions are predominately represented

with voxels of lower intensity. The visual contrast between voxels of the lesion and voxels of the spuriously highlighted regions decreases significantly in the SVM-G based confidence image. Similar characteristics of the presentation of lesions can be observed for the confidence volumes computed for case B015A.

Figure 5.12 presents coronal slices of the subtraction image and confidence volumes for the two malignant cases M007A and M015A. In the subtraction images, the lesions can be observed as conglomerates of voxels with moderate to high intensities. The regions for which the confidence images suggest suspicious tissue does not fill the entire contour of the lesion segmentation, but matches the region exposing a strong signal enhancement in the corresponding subtraction images. The evaluation of the image data of case M094A using the SVM-L leads to a confidence image, in which the lesion is displayed with high contrast to fat tissue. However, the glandular tissue surrounding the lesion exposes similar intensity values suggesting an extent of the disorder which is larger than the one described by the expert label. The confidence image based on the SVM-G provides a display of the lesion, in which the lesion is also silhouetted against glandular tissue.

Figure 5.13 illustrates the influence of the correction of a-priori probabilities on the initial visualisation of the confidence volumes. The top row shows a coronal slice of the two confidence volumes for case M094A and M007A based on the application of the SVM with uncorrected a-priori probabilities. The bottom row shows the same slices after the correction step. The assumption that suspicious examples are observed with the same a-priori probability as normal signals causes large areas of tissue exhibiting unclear temporal kinetic signals to be displayed with the similar intensity values as the lesion. After the correction step, voxels reflecting confidence values ranging from low to medium are visualised with lower intensities than before, whereas lesion voxels with high confidence values still expose high intensity values.

The performance of the machine learning based setup can further be increased by a preprocessing of the image data using a filter. Application of a mean filter of size $3 \times 3 \times 1$ to the coronal slices of the DCE-MRI data volumes increases the performance of the SVM-G evaluating the *raw*-feature to $\mathrm{Az} = 0.9657 \pm 0.0206$ and $\mathrm{Az}_{(\mathrm{SE}>0.9)} = 08076 \pm 0.0852$. In particular the $\mathrm{Az}_{(\mathrm{SE}>0.9)}$ value indicates an increased sensitivity. Figure 5.14 depicts the same coronal slices of the confidence volumes of case M007A (left) and M094A (right) as figure 5.12. The voxel intensities reflect the confidence values returned by the SVM-G evaluating temporal kinetic signals (*raw*-feature) of the filtered image volume. Now, the contour of the lesion in case M007A is nearly filled with bright voxels, whereas filtering predominantly leads to a reduction of spuriously highlighted regions in the confidence image of case M094A. Nevertheless, application of a filter or feature descriptions of larger image patches is always critical, if small lesion are to be detected. In this case, the temporal kinetic signals or features of normal tissue voxels are confounded with those of lesion voxels leading to an indistinct signal or feature vector.

The two images in the top row of figure 5.15 depict the same coronal slice of the subtraction image of case B013A based on the third postcontrast image. In the left image, voxel intensities are scaled to the range of $[0; 255]$ by suitable scaling and shifting factors. The right image presents the same slice but the values of the subtraction image are mapped to gray values by a manually tuned transfer function. While no signs of a suspicious signal enhancement can be observed in the former image, the latter indicates slightly and diffusely enhancing tissue in the lesion contour. The ROC analysis of the entire subtraction image returns an Az of 0.5924. The two images in the bottom row show the same coronal slice extracted from the confidence volume computed by the

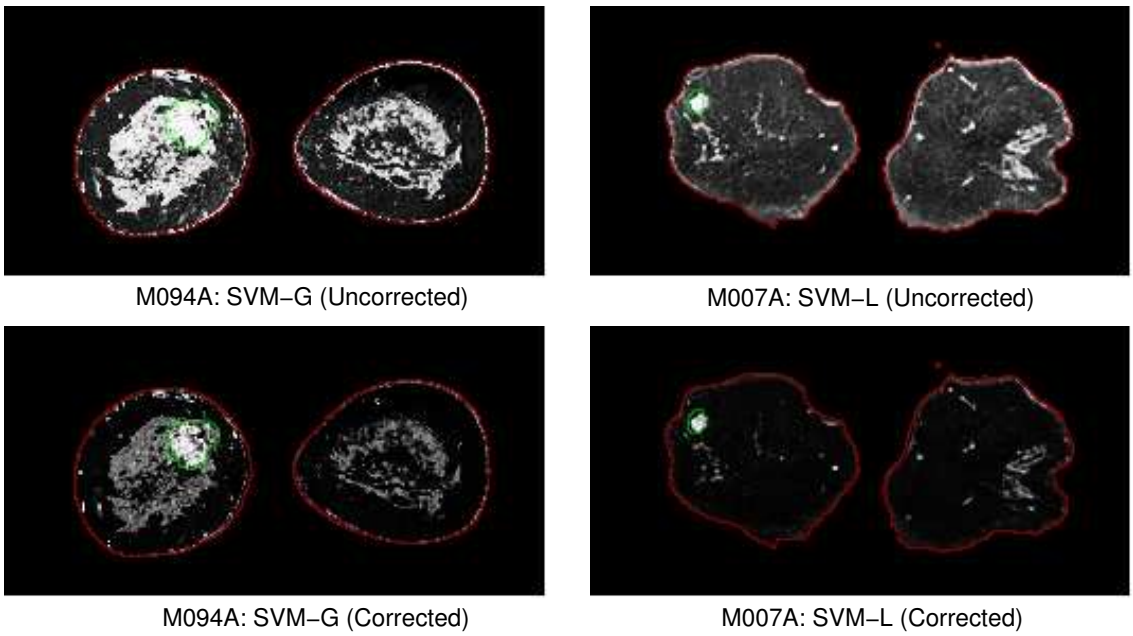


Figure 5.13: Coronal slices of two confidence volumes based on the SVM with and without corrected a-priori probabilities.

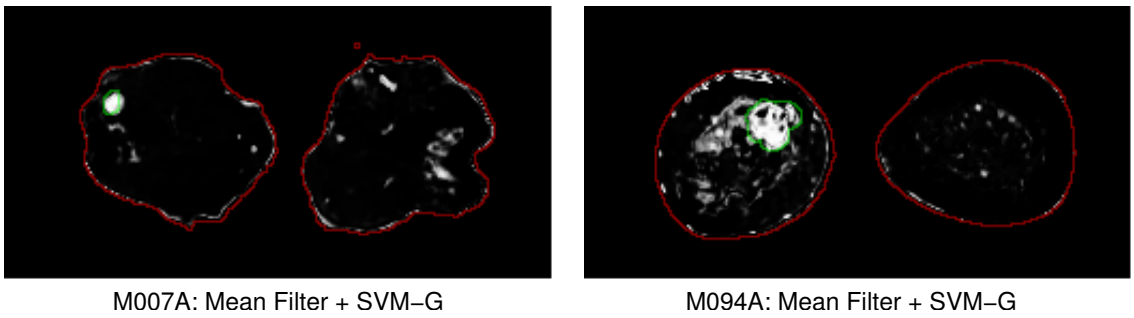
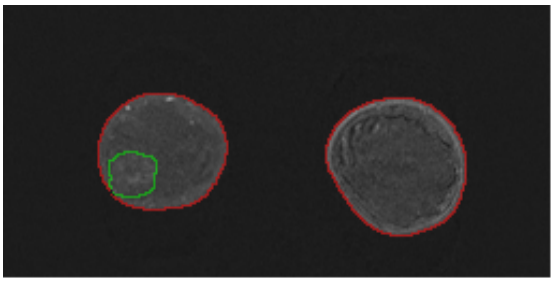
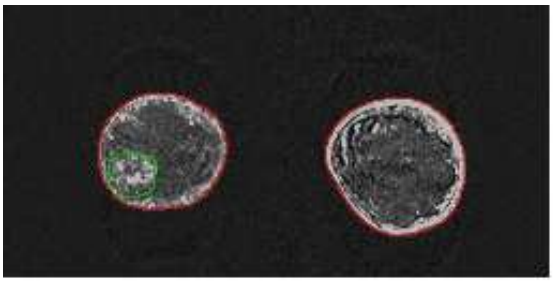


Figure 5.14: Coronal slices of the confidence volumes for the cases M007A and M094A of the *MARIBS* group based on the SVM-G evaluating the *raw*-feature. The training data and the test data is preprocessed by the application of a $3 \times 3 \times 1$ mean filter.

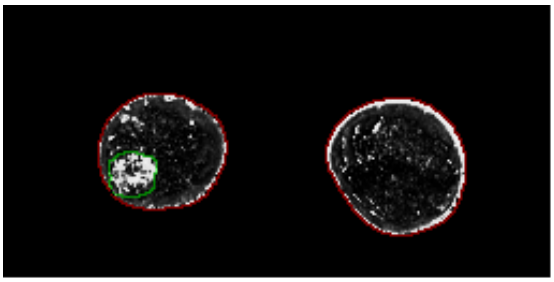
SVM-G evaluating the *allratio*-feature (left) and from the confidence volume based on the SVM-L evaluating the *raw*-feature. The mapping of the confidence values returned by the SVM to the gray values was also manually tuned to enhance the display of the benign lesion. The SVM-L based evaluation achieves an Az of 0.5912, whereas the SVM-G evaluating the allratios-feature achieves a value of 0.7978. Other coronal slices of all three image volumes display large regions of non-lesion tissue with bright voxels suggesting additional regions of suspicious tissue. The extent of the lesion tissue can not be recognised in terms of the signal enhancement of individual lesion voxels, but by considering the enhancement pattern and the form of a larger conglomerate of voxels displayed with heterogeneous intensity values.



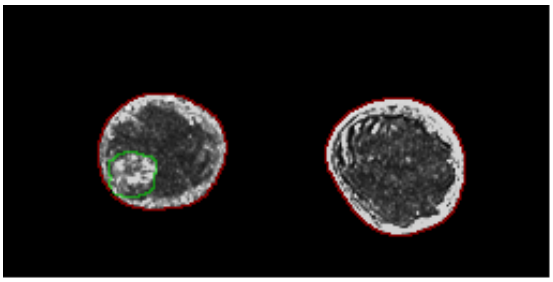
B013A: Subtraction Image



B013A: Tuned Subtraction Image



B013A: Tuned Confidence Volume (SVM–G)



B013A: Tuned Confidence Volume (SVM–L)

Figure 5.15: **Top row:** Coronal slice of the subtraction image for case B013A based on the third postcontrast image. In the left image, the voxel values of the confidence volume are scaled to the range of $[0; 255]$. In the right image, the mapping of voxel values to gray values is manually tuned in order the enhance the display of the lesion depicted by the green contour. **Bottom row:** Confidence image showing the same coronal slice as the subtraction image above. The left confidence image is based on the SVM-G evaluating the *allratio*-feature description of the temporal kinetic signal. The right image shows the same slice extracted from the confidence volume based on the SVM-L evaluating the unprocessed temporal kinetic signal. In both cases, the mapping of confidence values to gray values is manually tuned.

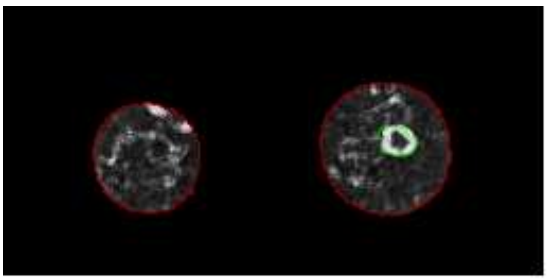
Region-Based Temporal Texture Features

Table 5.3 contains ROC indices Az and $\text{Az}_{(\text{SE}>0.9)}$ for the evaluation of the *MARIBS* group based on region-based temporal texture features. The highest values are obtained by the two linear classification algorithms SVM-L ($\text{Az} = 0.9207 \pm 0.0357$, $\text{Az}_{(\text{SE}>0.9)} 0.5949 \pm 0.1350$) and LDA ($\text{Az} = 0.9255 \pm 0.0306$, $\text{Az} = 0.5931 \pm 0.1578$). Similar to the *Munich* group, the performance of the detection setup based on texture features is inferior to the best setup based on voxel-based temporal kinetic features.

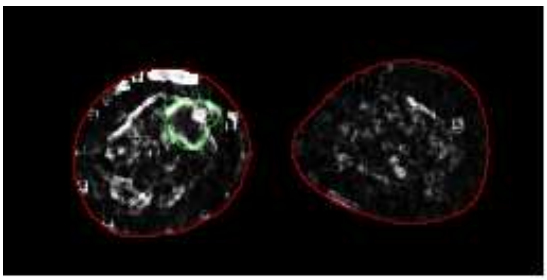
The reason for the inferior performance of the textural feature description can be exemplified on the basis of the two coronal slices of the confidence volumes B015A and M094A presented in figure 5.16. The intensity values reflect the outcome of the SVM-L evaluating Law's texture features. Voxels in the border region of the lesion are classified as suspicious with high confidence, whereas the evaluation of voxels of the lesion core leads to an assessment which suggests normal tissue. Since the textural features are calculated for subtraction images based on the first and last postcontrast images, the 5×5 windows of voxels in the border region are likely to display an edge between the region of normal tissue displayed with low intensities and the region of lesion tissue

Table 5.3: Mean values and standard deviations of Az and $Az_{(SE>0.9)}$ for the evaluation of the *MARIBS* group based on region-based textural features.

	Az	**$Az_{(SE>0.9)}$**
SVM-G	0.8735 ± 0.0536	0.4464 ± 0.2176
SVM-L	0.9207 ± 0.0357	0.5949 ± 0.1350
LDA	0.9255 ± 0.0306	0.5931 ± 0.1578



B015A: SVM–L + Texture Features



M094A: SVM–L + Texture Features

Figure 5.16: Coronal slices of the confidence volume for case B015A and M094A based on the evaluation of region-based temporal texture features using a SVM-L.

displayed with high intensities and therewith exhibit certain textural features. In contrast to this, the strongest distinguishing feature between the content of a window placed on the lesion core and a window placed on normal tissue is the average intensity level. In consideration of the fact that each Law's feature is normalised by the v^{L5L5} feature describing the average centre-weighted intensity level of the window content, both windows are likely to exhibit similar feature vectors and, therewith, do not allow to distinguish the underlying tissue by means of texture.

5.3 Discussion

The ROC analysis and the images displayed in the previous section have demonstrated how valuable machine learning techniques can be for the detection of tissue masses exhibiting suspicious temporal kinetic signals. The analysis of the subtraction images has shown that in particular subtraction images based on one of the later postcontrast images provide information for detecting suspicious tissue masses. Thus, in addition to the original image data depicting the anatomical structure of the entire breast, these subtraction images should be considered for an accurate and comprehensive evaluation of the DCE-MRI data.

All three classification algorithms are able to derive implicit models for the discrimination of temporal kinetic signals of normal and suspicious tissue, although the training cases are only partially labelled and no explicit a-priori knowledge about the signals of the tissue classes is used. For the regarded cases of the *Munich* group, ROC analysis indicates that the SVM with Linear Kernel provides sufficient discriminatory power to distinguish temporal kinetic signals of the different tissue classes. Regarding the examined feature transformations, the unprocessed temporal kinetic signal (*raw*-feature) is most suitable, although significance of the difference to the remaining feature could not be proven.

From the viewpoint of data visualisation, the confidence volumes obtained from the machine learning based approaches provide visualisations of the DCE-MRI data, which reflect an assessment of tissue based on the evaluation of the entire temporal signal information. The confidence volumes display tissue masses exhibiting suspicious temporal kinetic signals with high contrast to the surrounding normal tissue. According to the manual lesion segmentation of the radiologist, the highlighted tissue regions match the lesion masses to a high degree. Beside the improvement of the detection performance, the main value of the machine learning based approach arises from the facilitation of the manual data exploration process. The confidence volumes allow for localising lesion masses on the basis of examination of a single image volume per case which, for the *Munich* group of cases, does not require any further tuning of the data display.

For the *MARIBS* group, the SVM-G evaluating the *raw*-feature achieved the best results according to the ROC analysis. Considering the visualisations of the confidence volumes, the SVM-L provides a superior display of lesion masses for the benign cases, whereas the SVM-G displays more clearly the lesions of the malignant cases, in particular of case M094A. However, due to the low number of benign lesions in the *MARIBS* group it is questionable, whether this observation holds in general.

The textural description of small image patches has shown to be inappropriate for the detection of lesions. In particular for the cases of the *Munich* group, Law's texture features offered only inferior discriminative power in comparison with the voxel-based temporal kinetic features. An important disadvantage of textural features is that they always relate to larger image patches and are therewith inappropriate for precisely delineating the extent of small lesions. Lesions can exhibit fine structures such as spiculated extensions consisting of only a few voxels which are clearly smaller than the window for which textural features are computed. Therefore, the textural features are likely to refer to a mixture of normal and suspicious tissue. In contrast to the region-based temporal texture features, the voxel-by-voxel evaluation of the temporal kinetic signals allow for assessing the tissue characteristics with the spatial resolution of the image data, but is on the other hand more vulnerable to motion artifacts.

For a further discussion of the results, it has to be considered that the utilised lesion label is not a label reflecting the expert's assessment of the signal characteristics of individual temporal kinetic signals. The expert label marks those regions of tissue which after interpretation of the region's morphology, its temporal pattern of enhancement and its depiction by other imaging modalities give rise to the expert's final diagnosis. This assessment of tissue regions is likely to be different from the assessment of the temporal kinetic signals of the individual lesion voxels. In particular for the cases of the *MARIBS* group, several unlabelled subregions are indicated in the ANN based visualisation to be suspicious with moderate or high probability. Even though such regions are not rated as important for the final diagnosis, examination of the corresponding temporal kinetic signals reveals that the signals exhibit temporal kinetic signals which are indicative for pathological disorders. Furthermore, the contours of the lesion masses as segmented by the radiologist exceed the tissue masses displayed as suspicious for several cases of the *MARIBS* group. The temporal kinetic signals of such voxels which were marked as lesion voxels by the radiologist but designated as normal in the confidence volume frequently expose unsuspicious signal characteristics. Thus, from the viewpoint of evaluation of single kinetic patterns, the assessment of the ANN is reasonable. For the cases of the *Munich* group, the manual lesion segmentations widely conforms with the extent of tissue masses for which the ANN suggests

pathological disorders. This may point to the fact that labelling of such data is subjective and depends on the radiologist's expertise. The concordance between the contours of tissue masses displayed as suspicious by the ANN and the contours of the manual lesion segmentation may points to the radiologist's preference to segment lesions in the first instance by means of their temporal dynamic. In contrast to this, the comparison of the confidence volumes and manual lesion segmentations for the *MARIBS* group suggests that the radiologist also considered features such as the tissue morphology, which are not reflected by individual temporal kinetic signals. Thus, to obtain a more reliable assessment of the ANN performance, the label depicting the ground truth should be derived from more than one radiologist by e.g. computing the intersection of a repeated lesion segmentation by different radiologists.

In summary, the voxel-by-voxel evaluation of the DCE-MRI data using ANNs can not serve as a tool for fully automatic detection of lesions and can not entirely reproduce the radiologist's diagnosis. Nevertheless, the presented setup can serve as a tool that evaluates the entire temporal dimension of the multidimensional DCE-MRI data for computing visualisations, which enable the radiologist to localise and to delineate the extent of tissue regions which are suspected to exhibit a pathological disorder according to the corresponding temporal kinetic signals. In particular the confidence volumes computed for the cases of the *Munich* group display the lesions with high contrast to surrounding tissue, and the extent of the lesions can be recognised without further manipulation of the image display. The decision whether such tissue masses expose disorders with consequences for the final diagnosis demands for evaluating additional features which relate to the entire tissue mass such as the lesion's morphology or the distribution of benign and malignant signals inside the lesion.

6 Tissue Characterisation with Artificial Neural Networks

In the previous section, it was demonstrated how artificial neural networks and machine learning techniques can be applied to ease the localisation of temporal kinetic signals which are likely to be caused by benign or malignant tissue. The lesions of the considered cases were displayed as clusters of voxels with high intensity indicating suspicious characteristics of the associated signals.

In this chapter, the machine learning based analysis of DCE-MRI data will be extended towards the discrimination of temporal kinetic signals caused by malignant, benign and normal tissue. The trained predictor will provide a *pixel-mapping* which maps each temporal kinetic signal to a pseudo-colour reflecting characteristics of the course of the evaluated signal. The outcome will be a three-dimensional colour image of the breast which not only enables the radiologist to localise suspicious tissue masses but also to examine the distribution of benign and malignant signals in the lesion itself. Therewith, the technique provides the radiologist with valuable information for analysing and assessing lesions.

6.1 Motivation

In practice, the visual assessment of lesions by means of DCE-MR images is based on two complementary groups of features. *Morphological features* describe spatial properties of entire lesions. Spatial heterogeneity of enhancement of a lesion, irregularity of the lesion's contour or the characteristics of the lesion's margin are indicators for certain types of benign or malignant disorders [Heywang-Köbrunner and Beck, 1995]. For the examination of such growth patterns, single high-resolution contrast-enhanced MR images are most suitable. *Temporal kinetic features* describe the temporal behaviour of a tissue region after application of the contrast agent. Characteristics of the temporal course of a signal such as early or delayed enhancement, amount of enhancement or presence of a wash-out have demonstrated to be valuable features for discriminating malignant and benign tissue compartments. Due to the heterogeneity of lesion tissue, the assessment of tissue by means of its temporal behaviour demands for careful measurement of temporal kinetic signals. Analysis of the average signal of whole tumour ROIs, i.e. ROIs enclosing all voxels of lesions, have been commented by many authors to be inappropriate, in particular for evaluating malignant lesions where heterogenous areas of enhancement are diagnostically important [Collins and Padhani, 2004]. On the other hand, examination of kinetic signals of individual voxels is commonly too time-consuming since lesion masses frequently consist of several dozens to several hundreds of voxels. Furthermore, the assessment of a time curve associated with a single voxel has still to be correlated with the assessment of the time curves of the neighbouring voxels because isolated temporal kinetic patterns exhibiting a suspicious signal course may be found in any breast tissue due to the presence of larger vessels in the corresponding volume

element [Heywang-Köbrunner and Beck, 1995]. In consequence, temporal kinetic features are commonly analysed by means of average temporal signals of one or several ROIs which have been manually placed inside the lesion. Heywang-Köbrunner and Beck, 1995 stress that for an exact assessment of a lesion an attempt needs to be made to include exactly the areas of a lesion with the highest amount and speed of enhancement, whereas areas with lower amount and speed of enhancement must be excluded. Thus, the placement of ROIs strongly depends on the expertise of the radiologist and has been reported to be afflicted with a noticeable inter- and intra-observer variance [Orel and Schnall, 2001].

In order to facilitate the analysis of lesions and to attenuate the variability of the outcome, Collins and Padhani, 2004 demand that '*analysis and presentation of imaging data needs to take into account the heterogeneity of tumour vascular characteristics*'. A simple subtraction image reflects only a fraction of the temporal information of DCE-MRI data and the information is still fragmented and distributed over several images if all subtraction images based on the different postcontrast images are considered. Several authors propose the application of semiautomated methods such as pharmacokinetic models. Pixel-mapping techniques such as pharmacokinetic models expose several advantages such as appreciation of heterogeneity of enhancement and removal for the need to selectively place user-defined ROIs [Collins and Padhani, 2004], both reducing the risk of missing important information and the risk of creating ROIs containing different tissue types. However, pharmacokinetic models require considerable a-priori knowledge about the underlying physiological process and require DCE-MRI sequences consisting of a large number of postcontrast images to obtain reliable estimates for the model parameters. An alternative technique is calculating *parameter images* visualising estimates of e.g. the speed of signal enhancements by a suitable colouring of voxels. Nevertheless, a single parameter image only exploits a fraction of the information provided by the DCE-MRI data.

In this chapter, an adaptive pixel-mapping technique is proposed which is based on the application of artificial learning algorithms. This approach regards the temporal kinetic signals of several cases as a distribution of points in a signal space. The basic assumption of this approach is that, even though the tissue-specific distributions of normal, benign and malignant signals may considerably overlap, subregions exist in the signal space in which temporal kinetic signals of one tissue class prevail. In order to identify such regions, the algorithm analyses the distribution of the temporal kinetic signals of a small number of 'training cases'. The signals of these cases are attributed by a label either as normal, benign or malignant. This signal label is derived from a manual segmentation of the lesions and their subsequent histological examination. This combination of manual lesion segmentations with the outcome of the histological examination only provides an uncertain label for the temporal kinetic signals of individual lesion voxels, the outcome of the histological examination relates to the entire lesion mass. But in particular carcinoma consists of heterogeneous tissue such that the single label derived from the histological examination is not necessarily conform with the voxel label which would have been yielded by analysing the temporal kinetic signal associated with each individual lesion voxel. Nevertheless, the acquisition of both label components is frequently a part of the standard clinical diagnosis process. By training the ANN with the labelled data, the signal time courses are correlated with the given labels, and signal features that allow for distinguishing temporal kinetic signals caused by the three different tissue classes are autonomously derived from the data. Thus, implicit signal models are build by the ANN, and neither explicit knowledge about the underlying physiological

process nor the definition of an explicit mathematical model of this process or of the signals and their distinguishing features is necessary.

After the adaptation, the trained predictor can be applied to evaluate unlabelled temporal kinetic signals of unseen cases. The outcome of the predictor is a three-dimensional vector indicating the confidence that the evaluated signal was caused by one of the three tissue classes, respectively. This vector is used (i) for a classification of the signal as malignant, benign or normal and (ii) as a pseudo-colour reflecting the characteristics of the temporal kinetic signals. This pseudo-colour is determined by interpreting the vector as a RGB triplet which reflects the assessment of a signal such as *clearly malignant* or *border case between benign and malignant* by colour shadings. If the predictor is applied to all voxels of a DCE-MRI sequence, the outcome is a single RGB colour image, referred to as *scoring volume*. Alternatively, the application of the pixel-mapping can be limited to automatically selected subregions which are likely to contain tissue disorders. By means of the pseudo-colouring of entire image volumes or by visualisation of conventional DCE-MR images augmented by a pseudo-colouring of selected subregions, the radiologist is able to locate suspicious masses and to analyse the distribution of malignant and benign tissue inside potential lesions by visual examination of a single RGB image.

In summary, the proposed adaptive pixel-mapping technique extends the approach of parameter images in the way that it

- takes the entire temporal information of the kinetic signal into account,
- automatically derives features which are relevant for distinguishing signals of the three tissue classes and, therewith, needs no pre-defined features derived from explicit signal models for the assessment of temporal kinetic signals,
- provides a classification and characterisation of the temporal kinetic signals with respect to the considered tissue classes and not only a visualisation of quantitative measurements of pre-defined features,
- allows for visualisation of the outcome as a single colour image.

To demonstrate that the presented data-driven pixel-mapping technique provides a reasonable assessment of temporal kinetic signals, the outcome, i.e. the classification of individual voxels as malignant, benign or normal and the visual presentation of entire lesions, is compared with the outcome of the model-based *three-time-points* (3TP) technique. Although 3TP does not necessarily reflect the biological truth and can thus only provide an imperfect ground truth, it is a clinical standard technique for analysing DCE-MRI data in e.g. prostate and breast cancer diagnosis. The comparison with 3TP will illustrate that the machine learning based pixel-mapping provides an outcome which, despite the uncertain labelled training data, is widely concordant with the outcome of the model-based 3TP technique without demanding for a-priori knowledge or any model assumptions about the signal or the underlying biological process.

6.2 Data-Driven Pixel-Mapping Based on Supervised Learning

In the following, the different elements that constitute the data-driven adaptive pixel-mapping method will be described. Furthermore, detailed information about the training process, i.e.

about the acquisition of labelled training data and about adapting certain supervised learning algorithms, will be provided.

6.2.1 Setup for a Data-Driven Pixel-Mapping

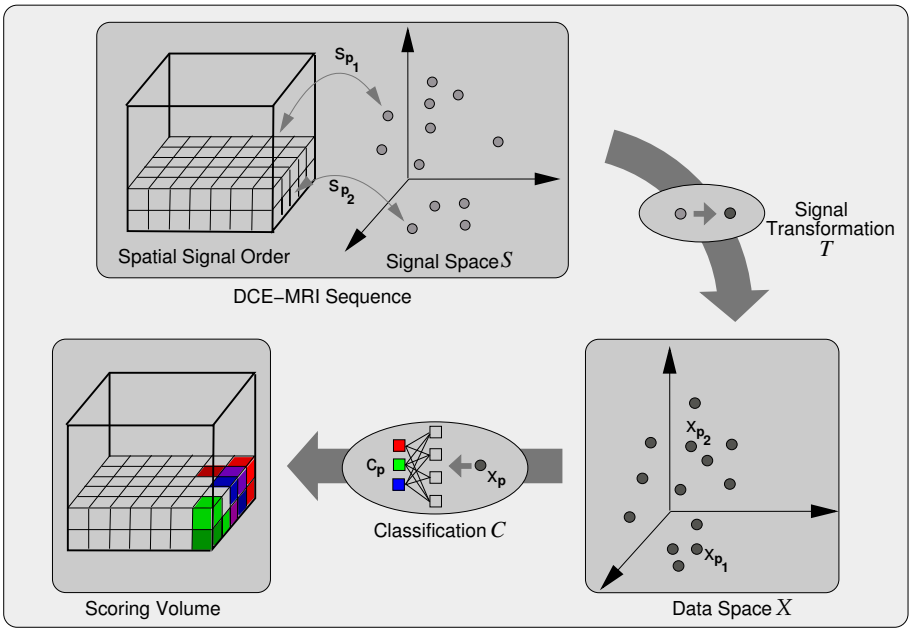


Figure 6.1: Illustration of the data-driven pixel-mapping based on machine learning algorithms. If the spatial order of the temporal kinetic signals is ignored, the signals $\mathbf{s} \in \mathcal{S}$ of several training cases form a data distribution in the signal space $\mathcal{S}$. The transformation T maps each signal $\mathbf{s_p}$ to a feature vector $\mathbf{x_p}$ providing a description of $\mathbf{s_p}$ which is suitable for discriminating the underlying tissue type. To this end, patterns $\mathbf{x_p}$ are evaluated by a supervised classification algorithm. The outcome of the classifier is a three-dimensional vector $\mathbf{y_p}$ which determines the RGB colour of the voxel $\mathbf{p}$ in the new three-dimensional scoring volume.

In order to visualise the DCE-MRI data as a single three-dimensional image augmented with information about the temporal course of kinetic signals, each signal $\mathbf{s} \in \mathcal{S}$ is evaluated by a pixel-mapping function

$$M : \mathcal{S} \to \text{RGB}, \quad \mathbf{s} \mapsto \mathbf{c}(\mathbf{s}) \tag{6.1}$$

based on supervised classification algorithms (see Fig. 6.1). The temporal kinetic signal $\mathbf{s_p}$ associated with the voxel $\mathbf{p}$ is mapped to a pseudo-colour $\mathbf{c}(\mathbf{s_p})$ (short $\mathbf{c_p}$) which reflects the probability that $\mathbf{s_p}$ was caused by malignant, benign or normal tissue. To this end, a transformation T maps each $\mathbf{s_p} \in \mathcal{S}$ to a feature vector $\mathbf{x}(\mathbf{s_p})$ (short $\mathbf{x_p}$), which is a point in a data space $\mathcal{X}$. A supervised classification algorithm is trained with a set of labelled data to distinguish between examples $\mathbf{x_p}$ of the $n_\Omega = 3$ tissue classes $\Omega = \{\text{Malignant}, \text{Normal}, \text{Benign}\}$. The classification algorithm is chosen such that the outcome of the discrimination function

$$C : \mathcal{X} \to [0;1]^{n_\Omega}, \quad \mathbf{x} \mapsto \mathbf{y}(\mathbf{x}) \tag{6.2}$$

is a n_Ω-dimensional vector $\mathbf{y}(\mathbf{x_p})$ (short: $\mathbf{y_p}$) with components $y_{\mathbf{p}k} \in [0;1], k = 1, \ldots, n_\Omega$. Hence, $\mathbf{y_p}$ can be directly mapped to the components of a RGB triple leading to the final pseudo-colour $\mathbf{c_p}$. Voxels classified with high confidence as malignant, normal or benign are coloured intense red, green and blue, respectively. Voxels with signal characteristics which do not allow a distinct assessment are displayed with corresponding colour shadings. Thus, suspicious masses are typically depicted as clusters of red, purple and blue voxels.

6.2.2 Preprocessing of Image Data

All images are preprocessed as already described in chapter 5. A binary mask, referred to as *breast mask*, is calculated for each image sequence using an adaptive threshold algorithm [Otsu, 1979]. The threshold is determined from the histogram of the precontrast image which was filtered using a three-dimensional median filter of size $3 \times 3 \times 3$. Application of a $5 \times 5 \times 5$ morphological closing operator to the outcome of the thresholding operation assures that the binary mask covers the image region of breast tissue without gaps. The breast mask of case m is formally described by the set $\mathrm{P}_m^{\mathsf{Breast}}$ of spatial coordinates $\mathbf{p}$ of voxels marked as breast tissue.

6.2.3 Preparing Training Data

For adapting the supervised classification algorithm, a set of training examples $\Gamma = \{(\mathbf{x}, \mathbf{y})_i\}, i = 1, \ldots, N$ consisting of pairs of input patterns $\mathbf{x}_i \in \mathbb{R}^{n_t}$ and class labels $\mathbf{y}_i$ is needed. The class label $\mathbf{y}_i$ encodes the membership of the corresponding input pattern $\mathbf{x}_i$ to one of the n_Ω classes using a *1-of-n_Ω* scheme, i.e. $\mathbf{y}_i = (1,0,0)$, $\mathbf{y}_i = (0,1,0)$ and $\mathbf{y}_i = (0,0,1)$ for examples of the malignant, normal and benign class, respectively. Samples $\Gamma^{\mathsf{Malignant}}$, Γ^{Benign} and Γ^{Normal} for the three tissue classes have to be selected from labelled DCE-MRI data volumes for constructing a representative set

$$\Gamma = \Gamma^{\mathsf{Normal}} \cup \Gamma^{\mathsf{Benign}} \cup \Gamma^{\mathsf{Malignant}} \tag{6.3}$$

subsequently used for adapting the classification algorithms.

Labelling of Image Data

The label $\mathbf{y_p}$ attributes each input pattern $\mathbf{x_p}$ either as a malignant, normal or benign example and is derived from two information sources which were acquired during a standard clinical DCE-MRI evaluation process. The spatial information about the location of lesion voxels is provided by a *manual lesion segmentation*. Voxels of lesions were manually marked by an experienced radiologist with a cursor on a screening device. To this end, the DCE-MRI data was presented as subtraction images visualising the temporal intensity gradient computed from one of the postcontrast and the precontrast image. Strong enhancing structures such as lesions or blood vessels appear with high intensity for a suitable selection of the postcontrast image. Furthermore, the radiologist correlated the DCE-MRI data with X-ray mammography images. Lesion voxels were then marked by either selecting individual voxels or by adjusting the vertex of a polygon enclosing a larger subset of voxels. The lesion segmentation of case m is formally described by the set $\mathrm{P}_m^{\mathsf{Lesion}}$ of spatial coordinates of designated lesion voxels.

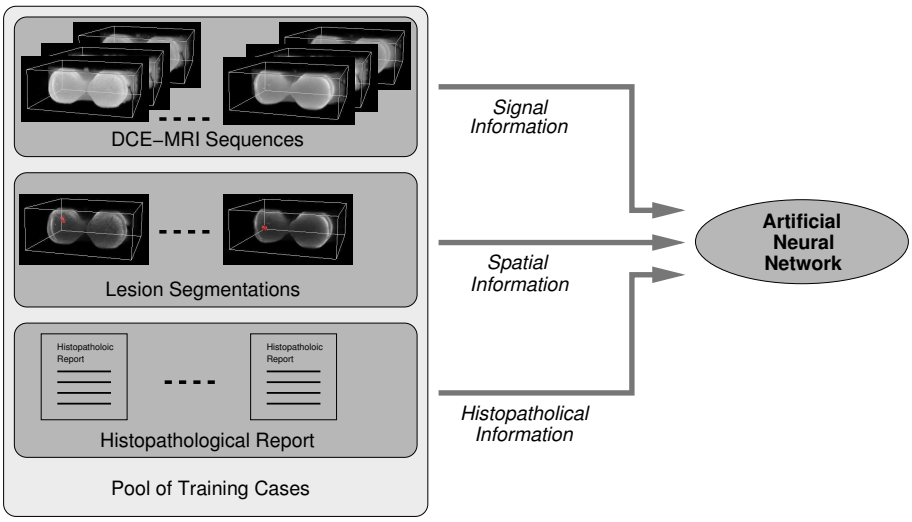


Figure 6.2: For the ANN adaptation, information from different sources is derived. The DCE-MRI sequences of a group of cases provide information about the signal domain. Spatial information about the corresponding lesions is derived from a manual lesion segmentation. A classification of the segmented lesions as malignant or benign is obtained from the laboratory report of the histopathological examination.

Even though the manual lesion segmentations yield information about the location of suspicious signals, no information about the distribution of benign and malignant signals inside the lesions is provided. Since lesion tissue is typically heterogeneous, a reliable classification of the temporal kinetic signals associated with individual lesion voxels requires a manual voxel-by-voxel evaluation by an experienced radiologist. Nevertheless, a voxel-by-voxel evaluation of a large number of lesions is impracticable under the prevailing circumstances of clinical diagnosis. An alternative source of information, albeit suboptimal for the purpose of preparing a set of labelled training signals, is the outcome of the histological examination of the lesion. The microscopic analysis of tissue samples extracted from a core-needle biopsy analyses the lesion tissue at the level of individual cells and allows for reliable classifications of type and grade of lesions. If the outcome of the histological examination is utilised as a label for temporal kinetic signals of lesion voxels, two aspects need to be considered:

- It is commonly not known which subsets of voxels exactly correspond to the tissue samples extracted for the histological examination. Thus, the histological report provides only a classification of the entire lesion without any further information about the exact location of malignant and benign signals inside the lesion. It has to be assumed that the distribution of benign and malignant training examples considerably overlap, if all signals of heterogeneous lesions are labelled according to the outcome of the corresponding histological examination.
- The histological examination evaluates tissue at a cellular level using features reflecting the type and configuration of individual cells. Consequently, the histological diagnosis does not necessarily reflect the diagnosis which a radiologist would have derived solely from the examination of the DCE-MRI data.

The advantage of signal labelling based on histological reports is the fact that histological examinations are often routinely performed for cases exhibiting suspicious lesions. Therewith, the DCE-MRI sequence and the corresponding histological reports are frequently available for a number of cases which is sufficient for adapting data-driven learning algorithms.

Training Data Selection

Despite the mentioned shortcomings of a signal label based on the histological examination, labelled examples for adapting the supervised learning algorithms are sampled by utilising the breast masks, the manual lesion segmentations and the outcomes of the corresponding histological examinations. Examples for temporal kinetic signals caused by malignant tissue are selected from the DCE-MRI data of the cases $M_m, m = 1, \ldots, n_M$ exhibiting lesions which were histologically classified as malignant:

$$\Gamma^{\mathsf{Malignant}} = \{(\mathbf{x}, \mathbf{y})_{\mathbf{p}} | \mathbf{p} \in \mathsf{P}_m^{\mathsf{Lesion}}, m \in \{M_1, \ldots, M_{n_M}\}, \mathbf{y} = (1, 0, 0)\}.$$

Examples representing signals caused by benign tissue are, respectively, selected from the cases $B_m, m = 1, \ldots, n_B$ exhibiting lesions classified as benign:

$$\Gamma^{\mathsf{Benign}} = \{(\mathbf{x}, \mathbf{y})_{\mathbf{p}} | \mathbf{p} \in \mathsf{P}_m^{\mathsf{Lesion}}, m \in \{B_1, \ldots, B_{n_B}\}, \mathbf{y} = (0, 0, 1)\}.$$

The set

$$\Gamma^{\mathsf{Normal}} = \left\{(\mathbf{x}, \mathbf{y})_{\mathbf{p}} | \mathbf{p} \in \mathsf{P}_m^{\mathsf{Breast}} \wedge \mathbf{p} \notin \mathsf{P}_m^{\mathsf{Lesion}}, m \in \{M_1, \ldots, M_{n_M}, B_1, \ldots, B_{n_B}\}, \mathbf{y} = (0, 1, 0)\right\}$$

of examples of normal tissue signals is selected from all positions marked by the breast mask, excluding the lesion segmentation. Subsequently, the three sets $\Gamma^{\mathsf{Malignant}}$,$\Gamma^{\mathsf{Benign}}$ and Γ^{Normal} are divided into two subsets which are used for adapting the learning algorithm and for selecting the algorithm's hyperparameters. The size of each subset is chosen under consideration of the computational expenses of the adaptation and hyperparameter selection steps and is described in more detail in section 6.2.4 and 6.2.5.

Feature Description

Examination of temporal kinetic patterns $\mathbf{s}$ as measured for small ROIs is a common method for characterising lesion masses. However, the categorisation of such signal time courses based on visual examination is insufficiently standardised [Szabo et al., 2003] and, therewith, is subjective and depends on the radiologist's expertise. Although none of the different interpretation strategies used in literature has evolved into a generally adopted approach, some basic features for quantitative evaluation of temporal kinetic signals are widely used. Szabo et al., 2003 examined the value of different morphologic and kinetic features for the formulation of a systematic scoring scheme for lesion characterisation. Among others, the following temporal kinetic features were examined:

- The *percentage enhancement* reflecting the increase of signal value in the n-th postcontrast image relative to the value in the precontrast image:

$$E_n(\mathbf{s}) = \frac{s_n - s_{\mathsf{pre}}}{s_{\mathsf{pre}}} \cdot 100 \tag{6.4}$$

- The *initial slope* reflecting the slope of the signal uptake between the precontrast value and the signal peak:

$$\text{Slope}(\mathbf{s}) = \frac{E_{\text{peak}}(\mathbf{s})}{T_{\text{peak}}(\mathbf{s})} \tag{6.5}$$

  with the maximum percentage enhancement $E_{\text{peak}}(\mathbf{s})$ and the time-to-peak $T_{\text{peak}}(\mathbf{s})$.

- The *washout ratio* reflecting the downslope from the signal's peak to it's value in the n-th postcontrast image ($n >$ peak):

$$W_{\text{peak}-n}(\mathbf{s}) = \frac{s_{\text{peak}} - s_n}{s_{\text{peak}}} \cdot 100 \tag{6.6}$$

Additional quantitative features of temporal kinetic signals have been examined by Abdolmaleki et al., 2001. The mentioned features can be computed for signals which relate to a single voxel as well as for averaged signals of ROIs. In the latter case, the spatial variance of the different features in the ROI can also be considered as it is done in the CAD system proposed by Chen et al., 2004. In general, the definition of features and the selection of reasonable thresholds for the discrimination of benign and malignant signals often depends on parameters of the DCE-MRI protocol such as the spatial and temporal resolution.

To emphasise the fundamental idea of a data-driven approach to tissue characterisation and to avoid explicitly defined quantitative features, two types of signal transformations

$$T : \mathcal{S} \to \mathcal{X}, \quad \mathbf{s} \mapsto \mathbf{x} \tag{6.7}$$

for the computation of a feature description $\mathbf{x}$ of the measured signal $\mathbf{s}$ are investigated. The first feature description, referred to as the *raw*-feature, consists of the unprocessed signal values. The second feature, referred to as the *allratio*-feature, describes the temporal course of signal intensity as the ratios of intensity values at two different points in time:

$$\mathbf{x} = \left(\frac{s_j}{s_k}\right) \tag{6.8}$$

with $j, k = 1, \ldots, n_t$ and $k < j$. Both features have already been employed in chapter 5 for the detection of suspicious tissue masses.

6.2.4 Adaptation of Multiclass Support Vector Machines

Support vector machines (SVM) have been applied to various classification tasks and have become the state-of-the-art algorithm due to their superior classification performance. Thereby, multiclass tasks have to be decomposed into a sequence of binary classification tasks, since SVMs were originally only designed for problems of the latter type. Since the given task requires discriminating temporal kinetic signals of three different classes, it is decomposed into a sequence of binary subtasks following a *one-vs-all* scheme (see chapter 4.2.3). The k-th subtask, $k = 1, \ldots, n_\Omega$, is solved by a binary SVM which, after adaptation, returns a margin value $m_k(\mathbf{x}) = f_{\boldsymbol{\theta}_k}(\mathbf{x})$ for each

Table 6.1: Ranges of values for the regularisation parameter C and kernel parameter σ which are scanned for choosing a suitable MSVM hyperparameters.

Kernel Type	Regularisation Parameter	Kernel Parameter
Linear Kernel	$C \in \{10^{\frac{2i}{10}}\}, i = 1, \ldots, 10$	none
Gaussian Kernel	$C \in \{10^{\frac{2i}{10}}\}, i = 1, \ldots, 10$	$\sigma = 0.4, 0.8, \ldots, 4$

evaluated example. Since each tissue class is regarded by one of the binary subtasks as the target class to be distinguished from the union of the remaining classes, the outcome of the *multiclass SVM* (MSVM) is a n_Ω-dimensional vector $\mathbf{m}(\mathbf{x})$ of margin values. The class label assigned by the SVM to the evaluated example is indicated by the index of the vector component with the largest value. In order to map the margin vector $\mathbf{m}(\mathbf{x})$ to the final output vector $\mathbf{y}(\mathbf{x})$ with components $y_k(\mathbf{x}) \in [0; 1]$, the output of the MSVM is postprocessed by a *softmax* function

$$y_k(\mathbf{x}) = \frac{\exp(a_k m_k(\mathbf{x}) + b_k)}{\sum_i^{n_\Omega} \exp(a_i m_i(\mathbf{x}) + b_i)} \tag{6.9}$$

with parameters $a_k, b_k,\ k = 1, \ldots, n_\Omega$ which have to be adapted during training of the MSVM (cf. 4.2.5).

Adaptation of the MSVM algorithm requires the manual selection of the hyperparameter C controlling the regularisation of the learning process and of the hyperparameters of the kernel function. Suitable hyperparameters are separately selected for both types of signal transformation and kernel function by a grid search in the corresponding parameter spaces. To this end, the performance of the MSVM trained with examples from the training set Γ_{Train} is evaluated on a test set Γ_{Test} for different combinations of hyperparameter values. The training set Γ_{Train} consists of 50% of the set $\Gamma^{\text{Malignant}}$ of malignant examples, 50% of the set Γ^{Benign} of benign examples and an equal number of normal signals randomly selected from Γ^{Normal}. All three sets $\Gamma^{\text{Malignant}}$, Γ^{Benign} and Γ^{Normal} only contain data sampled from the training cases. The remaining examples of $\Gamma^{\text{Malignant}}$ and Γ^{Benign} and a tenfold number of normal signals are used for testing. Each input component of the examples in Γ_{Train} and Γ_{Test} is normalised to zero mean and unit variance. The MSVM is employed with the linear kernel (MSVM-L) and the nonlinear Gaussian kernel (MSVM-G) leading to a one-dimensional and two-dimensional hyperparameter space, respectively. During the grid search in the hyperparameter space, the MSVM performance is assessed for the parameter combinations as outlined in table 6.1.

Training of the MSVM for a certain choice of kernel function and hyperparameters consists of the following steps:

1. A three-fold cross validation on the training set in order to obtain an unbiased margin vector $\mathbf{m}(\mathbf{x}_i)$ for each of the N examples $(\mathbf{x}, \mathbf{y})_i \in \Gamma_{\text{Train}}$.

2. Training of the MSVM with the entire training set.

3. Adaptation of the softmax function (6.9) using the margin vectors computed in step 1. The

parameters are adapted by minimising the *cross-entropy* error

$$E = -\sum_{i=1}^{N}\sum_{k=1}^{n_\Omega} y_{ik}\log(y_k(\mathbf{x}_i)) \tag{6.10}$$

for the Γ_{Train} using the *Polak-Ribière* variant of the conjugate gradient method [Press et al., 1986].

The performance of the trained predictor is subsequently estimated on the test set Γ_{Test} by measuring the balanced accuracy, i.e. the mean accuracy of the three classes. Those hyperparameters yielding the highest performance value are assumed to be the optimum for the regarded type of feature and the corresponding predictor is employed for the evaluation of unseen cases.

6.2.5 Adaptation of Local Sigmoid Maps

Like the *local linear map*, the *local sigmoid map* can rapidly be trained even for huge data set due to the iterative nature of the algorithm. Therefore, the entire set of examples Γ sampled from the training cases is utilised for the training process. Since normal examples outnumber malignant and benign examples by far, the input patterns for the adaptation are selected from Γ^{Normal}, $\Gamma^{\text{Malignant}}$ and Γ^{Benign} with a predefined probability of $P(\text{Normal}) = 0.8$, $P(\text{Malignant}) = 0.1$ and $P(\text{Benign}) = 0.1$.

The parametrisation of the time-dependent learning rates and the neighbourhood function have demonstrated to be uncritical and are set to a fixed setting throughout the experiment. For the vector quantisation step, the *neural gas* algorithm is employed with a learning rate α_t linearly decreasing from $\alpha_0 = 0.5$ to $\alpha_{N_t} = 10^{-4}$. The initial value of the width of the neighbourhood function $\sigma_0 = \frac{n_p}{2}$ is chosen according to the number of prototypes n_p and linearly decreases to the final value of $\sigma_{N_t} = 10^{-2}$. The prototypes are initialised with randomly selected examples from the training data.

The learning rate ϵ_t of the local experts decreases linearly from an initial value of $\epsilon_0 = 0.25$ to $\epsilon_{N_t} = 10^{-4}$. The weights $\boldsymbol{\beta}$ of the local experts are initialised with random values following a normal distribution $\mathcal{N}(\mu, \sigma)$ with $\mu = 0$ and $\sigma = 1$.

Number of Prototypes

The complexity of the discrimination function of the LSM is controlled by two hyperparameters. The number of prototypes n_p defines the number of subregions the data space is divided into and, therewith, how many local experts implementing linear discriminations functions constitute the piecewise linear discrimination function of the entire LSM architecture. The second hyperparameter is the smoothing parameter ϕ controlling smoothness of the discrimination function at the conjunction of two neighbouring local experts, i.e. linear pieces of the discrimination function.

The optimal choice of the two hyperparameters is separately selected for each feature type by a grid-search. To this end, a disjoint test set Γ_{Test} is isolated from the training data $\Gamma_{\text{Train}} = \Gamma$ containing 50% of the malignant and benign signals and ten times as much examples of normal signals. For each parameter tuple (n_p, ϕ) with $n_p \in \{1, 2, \ldots, 15\}$ and $\phi = \{0, 0.05, 0.1, \ldots, 1.0\}$, the performance of the LSM is estimated by measuring the balanced accuracy on the test set.

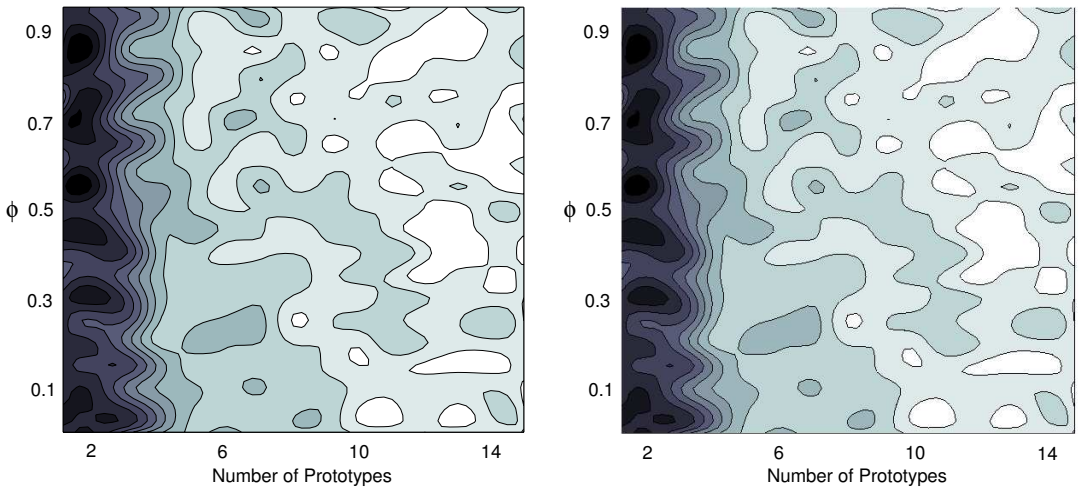


Figure 6.3: Isocontour plot of the balanced accuracy for different parameterisations of the number of prototypes n_p (x-axis) and the smoothing parameter ϕ (y-axis). The left plot depicts the balanced accuracy values for the LSM evaluating the *raw*-features. The right plot reflects the balanced accuracy computed with the LSM evaluating the *allratio*-features.

Figure 6.3 depicts two isocontour plots of the balanced accuracy for different values of (n_p, ϕ). For the *raw*-features, high accuracy values are obtained within the range of $n_p = \{3, 4, \ldots, 6\}$ and $\phi = [0; 1]$. The performance clearly decreases for large numbers of prototypes. According to the test error, the most suitable parameterisation is $n_p = 4$ and $\phi = 0.4$. In the case of the *allratio*-features, accuracy increases with the number of prototypes and the lowest error is yielded for $n_p = 13$ and $\phi = 0.4$.

6.2.6 Evaluation

The evaluation of the setup is based on the evaluation of the DCE-MRI sequences as recorded for cases of the *Munich* group, which consists of a balanced number of cases with histologically proven malignant and benign lesions. The outcomes of the LSM and MSVM based pixel-mapping are evaluated following a *leave-one-case-out* scheme. Each algorithm is solely adapted only with training data sampled from eleven of twelve cases. The remaining case is used to evaluate how well the method performs on a new unseen case. This evaluation scheme is repeated twelve times, each time using a different case for validation. In order to attenuate the influence of random initialisation and random selection of training data, the entire scheme is repeated five times. The five scoring volumes obtained from the repetitions of the experiment for each validation case are subsequently averaged yielding a single scoring volume for each validation case.

6.3 Results

All images and quantitative performance measures presented in this section refer to the averaged scoring volumes. The section is subdivided into two parts. The first part addresses the visualisation

of entire scoring volumes and the display of conventional DCE-MR images augmented by a pseudo-colouring of automatically selected subregions which are likely to represent suspicious tissues.

The second part is focused on the visualisation and the assessment of lesion masses. The outcomes of the LSM and MSVM based pixel-mappings are compared with the results obtained from the 3TP technique by means of images of pseudo-coloured lesion masses. Furthermore, to quantify the concordance between the data-driven approaches and the model-based 3TP technique, a voxel label is determined using 3TP. This label attributes each voxel of the lesion either as benign, malignant or suspicious with indistinct signal characteristics and serves as gold standard for a quantitative evaluation. Although 3TP is an imperfect gold standard which does not necessarily reflect the biological truth, a comparison with this clinically applied standard technique allows for assessment of the reasonability of the ANN outcome.

6.3.1 Visualisation of Entire Image Volumes

Pseudo-Colouring of Entire DCE-MR Images

Figure 6.4 depicts axial slices of the scoring volume of the two benign cases B1 and B2 and the two malignant cases M4 and M6. The kinetic signals $\mathbf{s_p}$ of voxels marked by the breast mask were evaluated with the LSM using the *raw*-features. The outcome $\mathbf{y_p}$ reflecting the confidence that the signal belongs to the class of malignant ($y_{\mathbf{p}1}$), normal ($y_{\mathbf{p}2}$) and benign ($y_{\mathbf{p}3}$) was mapped to the red, green and blue component of the pseudo-colour $\mathbf{c_p}$, respectively. All four lesions are easy to locate in terms of their size and pseudo-colours assigned to the lesion voxels. The lesion masses of B1 and B2, which are both attributed as benign according to the histological examination, are displayed by voxels with shadings of blue to purple indicating high confidence values for the benign and slightly increased values for the malignant component of the LSM output. The malignant lesion of case M4 exposes a ring of intense red and purple voxels surrounding a green core which may consist of necrotic tissue with low vascularity. The average signal of the whole-tumour ROI exhibits a delayed uptake and weak washout between the fourth and fifth postcontrast image suggesting a benign disorder of the tissue. In this situation, the pseudo-colours provide valuable architectural information about the lesion mass and allow for localising subregions which are malignant at high probability. The lesion of case M6 is a bifocal lesion consisting of a benign (intense blue) and a malignant (intense red) compartment. The visualisations of the lesions of all twelve cases will be investigated in more detail at the end of this section. The major part of the breasts is homogenously depicted with intense green voxels indicating breast tissue which is healthy at high probability.

Selective Augmentation of Conventional DCE-MR Images

An alternative visualisation of the data can be obtained by superimposing pseudo-colours only onto those voxels whose signals are rated as suspicious with a probability above a certain threshold. The depiction of the entire DCE-MRI volume with pseudo-colours facilitates the localisation and examination of suspicious masses due to the high visual contrast between the lesion voxels depicted with shadings of blue and red and the surrounding normal tissue homogenously coloured green. In contrast, a limited augmentation of conventional DCE-MR images with pseudo-colours enables

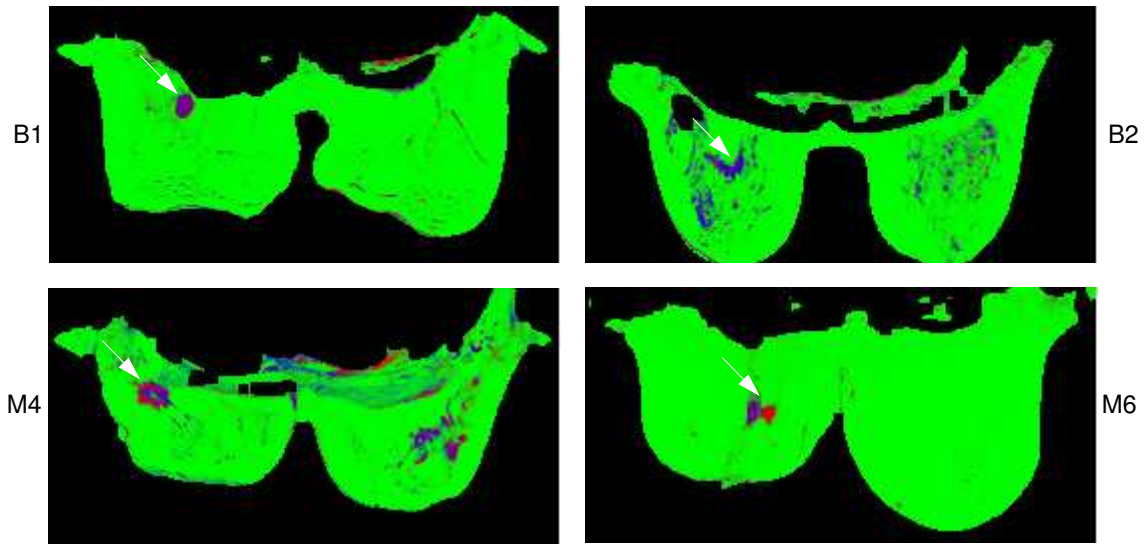


Figure 6.4: Subregion of an axial slice of the LSM based scoring volume for the benign cases B1 and B2 (top row) and the malignant cases M4 and M6 (bottom row). The locations of lesions are designated by white arrows. In all four images, the major part of the breast consisting of normal tissues such as fat or glands is depicted intense green. The benign lesions are presented as intense blue or purple masses, whereas the carcinoma of the malignant cases are predominately displayed with intense red and purple voxels. Next to the lesion masses, the pseudo-colouring suggests additional benign masses for case B2 (right breast, below the lesion) and M4 (left breast, centre).

the observer to interpret the spatial distribution of the pseudo-colours in regions of suspicious tissue in the context of the structure of the surrounding tissue.

For an automatic selection of regions which are to be presented in pseudo-colours, the lesion detection setup as proposed in the previous chapter can be applied. The voxel value in the confidence volume computed for the patient under investigation reflects the probability $P(\text{Suspicious}|\mathbf{x_p}) = P(\neg\text{Normal}|\mathbf{x_p})$ that the associated temporal kinetic signal was caused by suspicious tissue. A comparable confidence value can be obtained from the outcome $\mathbf{y}$ of the LSM and MSVM, respectively. The second component $y_{\mathbf{p}2} = P(\text{Normal}|\mathbf{x_p})$ reflects the probability that the temporal kinetic signal associated with voxel $\mathbf{p}$ is caused by normal tissue. Hence, a probability value $P(\neg\text{Normal}|\mathbf{x}) = 1 - P(\text{Normal}|\mathbf{x_p}) = 1 - y_{\mathbf{p}2}$ above a certain threshold indicates suspicious temporal kinetic signals.

Examples for selectively augmented conventional DCE-MR images can be observed in figure 6.5. The figure depicts the same axial slices as figure 6.4, but pseudo-colours are only superimposed onto voxels $\mathbf{p}$ with $P(\neg\text{Normal}|\mathbf{x_p}) > 0.5$. Thereby, tissue regions with an increased probability of being malignant or benign are presented in pseudo-colours, whereas regions depicting glands, fat or muscle tissue exposing unsuspicious kinetic signals are displayed with the gray values of the precontrast image. Through this, suspicious regions are augmented by information about the temporal data component which therefore be correlated with the anatomical structures of the surrounding tissue.

In order to evaluate whether $P(\neg\text{Normal}|\mathbf{x_p}) = 1 - y_{\mathbf{p}2}$ is a reasonable criterion for the selection

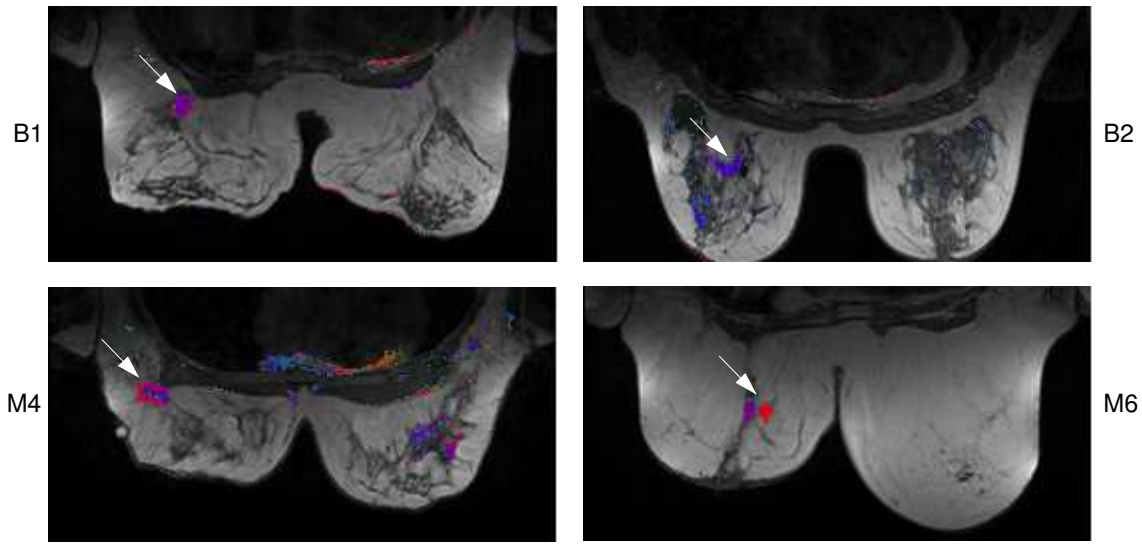


Figure 6.5: Subregion of an axial slice of the LSM based scoring volume for the benign cases B1 and B2 (top row) and malignant cases M4 and M6 (bottom row). The figure depicts the same slices as figure 6.4, but the pseudo-colouring is only superimposed onto voxels with an increased probability of suspicious signal characteristics according to $P(\neg\text{Normal}\mathbf{x_p}) > 0.5$. The remaining voxels reflect the intensity value of the precontrast image.

Table 6.2: Az (first and second column)and $Az_{(SE>0.9)}$ (third and fourth column) for the different combinations of classification algorithms and feature types.

	Raw	**Allratios**	**Raw**	**Allratios**
LSM	0.9912	0.9825	0.9667	0.9112
MSVM-L	0.9905	0.9865	0.9658	0.9689
MSVM-G	0.9899	0.9880	0.9535	0.9499

of suspicious subregions, a ROC analysis is performed. Similar to the ROC analysis in the previous chapter, the indices *area-under-the-ROC-curve* (Az) and *partial-area-under-the-ROC-curve above a sensitivity of 0.9* ($Az_{(SE>0.9)}$) are calculated. The ROC curve for each combination of classifier (LSM/MSVM-L/MSVM-G) and feature (*raw*/*allratio*) is calculated by *pooling* of the validation data [Swets and Pickett, 1982]: The data of the twelve average scoring volumes are merged (*pooled*) and a single ROC curve is computed using the manual lesion segmentation as ground truth. Subsequently, the indices Az and $Az_{(SE>0.9)}$ are determined for each ROC curve. The results are listed in table 6.2. Classifiers trained with the *raw*-features outperform classifiers trained with *allratio*-features. Only in the case of the MSVM-L, a slightly higher $Az_{(SE>0.9)}$ is yielded for the *allratio*-features.

Spuriously Highlighted Tissue Regions

For the cases B2 and M4, the visualisation points to additional regions of suspicious tissue which were not marked by the radiologist during the manual segmentation of the lesion. For instance in

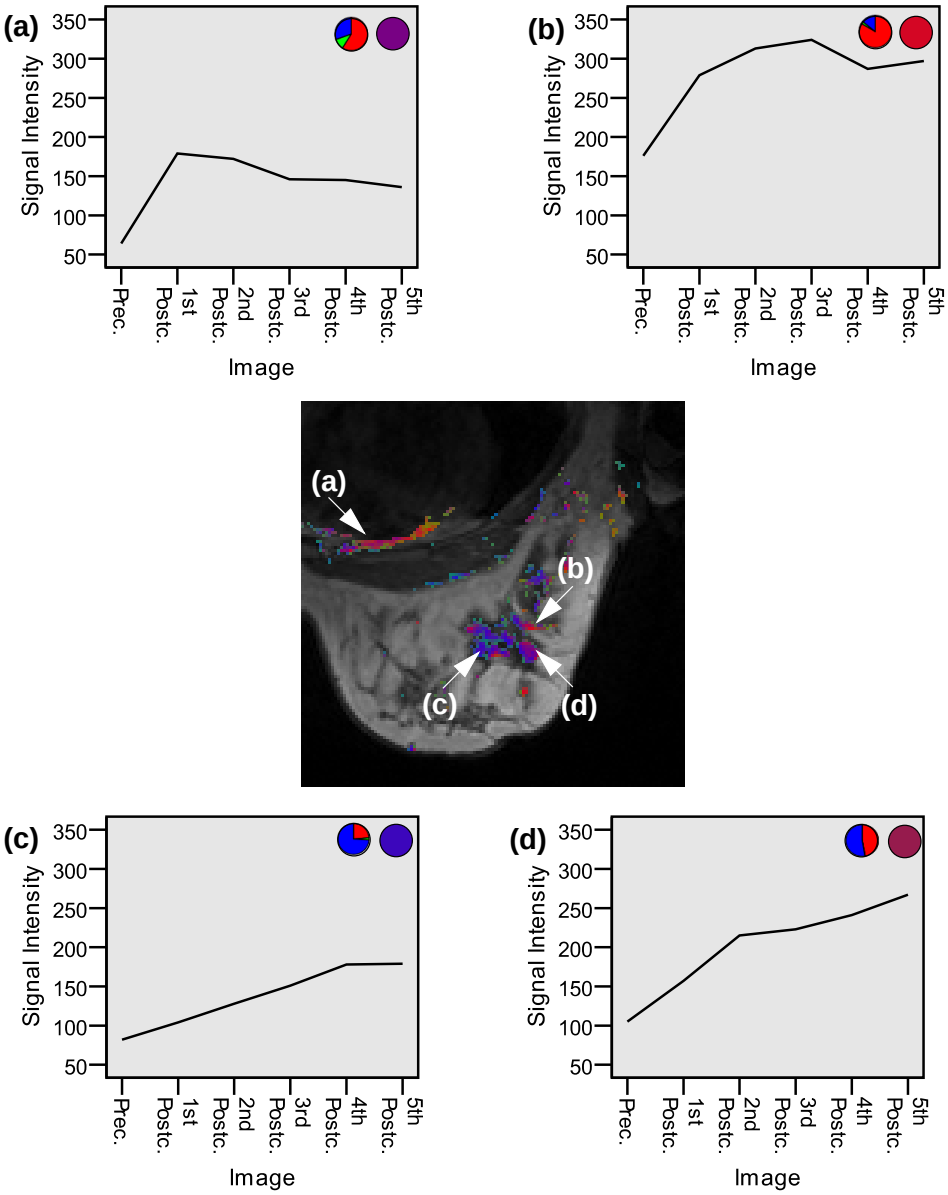


Figure 6.6: Pseudo-colouring superimposed onto a subregion of the precontrast image of the left breast of case M4. The image exhibits several regions of tissue for which the pseudo-colouring indicates suspicious signal characteristics ($P(\neg\text{Normal}|\mathbf{x_p}) > 0.5$), but which are false positives according to the expert label. Though, exemplary examination of the kinetic signals of such false positive voxels as can be observed on the plots affirms that the kinetic signals exhibit characteristics which are indicative for benign and malignant tissue.

figure 6.6, the pseudo-colouring superimposed onto the precontrast image of the left breast of case M4 highlights several regions by pseudo-colours indicating malignant or benign tissue. According to the expert label, the lesion of case M4 is located in the right breast. Hence pseudo-coloured

regions in the left breast have to be treated as spuriously highlighted tissue masses. Though, examination of the temporal kinetic signals of such spuriously highlighted regions reveals that, from the viewpoint of assessment of temporal kinetic signals of single voxels, the outcome of the pixel-mapping appears reasonable. The four plots in figure 6.6 visualise temporal kinetic patterns as measured inside clusters of spuriously highlighted voxels. Additionally, a pie chart illustrates the distribution of the 'probability mass' onto the three RGB components of the pseudo-colour. The pseudo-colour itself is reflected by a filled circle to the right of the pie chart. Plot (a) exposes a kinetic signal with an uptake in the early postcontrast period followed by a steady wash-out and, therewith, exposes a typical course of a malignant kinetic signal. Correspondingly, the LSM classifies the signal as malignant with a probability which is twice the probability of a benign classification. Since the region is located inside the thorax, it can be easily distinguished from potential lesion masses. The signal shown in plot (b) is classified with high confidence as malignant. The signal exhibits a delayed but strong uptake followed by a weak wash-out in the very late postcontrast period. A signal which steadily increases over the entire period of time can be observed in plot (c). The signal is classified by the LSM as being benign leading to an intense blue pseudo-colour. The last plot depicts a steadily increasing signal with a fast uptake during the first 220sec, followed by a period of slower signal enhancement. The assessment by the LSM indicates that the signal is more likely to be caused by benign tissue than by malignant tissue.

6.3.2 Visualisation of Lesion Masses

Analysing the Relations between Pseudo-Colours and Temporal Kinetic Signals

In order to analyse the data-driven pixel-mapping function, the pseudo-colours assigned to lesion voxels and the corresponding temporal kinetic signals have to be examined simultaneously. Only a simultaneous presentation of both types of information allows for detailed examination of the relationship between the pseudo-colour and the temporal kinetic signal. Since this procedure is impracticable for all lesion voxels, the relation is exemplarily analysed by means of a single lesion and a small number of temporal kinetic signals.

Figure 6.7 depicts a subregion of the precontrast image recorded for case M1. Lesion voxels as marked by the radiologist are coloured with pseudo-colours obtained from the data-driven pixel-mapping based on the LSM (*raw*-features). The plot presents temporal kinetic signals as measured at three different positions inside the ductal carcinoma. The lesion mass can be divided into three subregions. The upper right part of the lesion is coloured with shadings of blue to purple characterising the underlying tissue as benign. Signal (a) was measured at the position of an intense blue voxel. The signal exposes typical properties of a benign signal with a slow, sustained enhancement during the major part of the considered space of time followed by a late wash-out. The middle part of the lesion is coloured with shadings of red indicating a malignant type of tissue. The temporal kinetic signal (b) as measured at the position of an intense red voxel exhibits a signal course which is characteristic for tissue affected by malignant disorders. The lower part of the lesion exhibits a spiculated margin with two extensions. These extensions are characterised by the pseudo-colouring as normal tissue. The signal (c) as measured for one of such green coloured voxels reveals that the kinetic pattern exhibits indistinct signal characteristics.

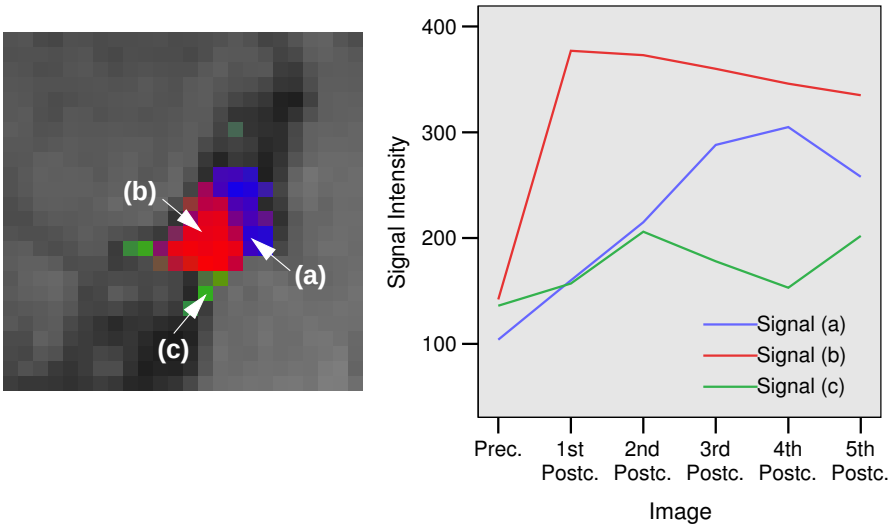


Figure 6.7: Visualisation of the lesion of case M1 based on LSM (*raw*-features). Additionally, three kinetic signals corresponding to voxels which are characterised as benign (a), malignant (b) and normal (c) are depicted.

3TP as a Model-Based Reference Technique

In order to analyse the presentation of the remaining lesions, pseudo-colourings of the lesions based on the different classification algorithms are compared to those obtained from the 3TP technique. Since the pseudo-colours red and blue derived from the data-driven pixel-mapping techniques and 3TP expose a comparable semantic meaning, the comparison of the lesion visualisations allow for appreciation, whether the data-driven approach provides a reasonable assessment of the signals in terms of congruence with a standard model-based diagnosis technique. The green colour has different meanings in the two types of pseudo-colouring techniques. In case of 3TP, intense green indicates a strong signal uptake between the precontrast and the first postcontrast image followed by indistinct wash-out characteristics. The corresponding signals are likely to be suspicious due to the strong uptake, but cannot be attributed as malignant or benign by means of the signal course in the late postcontrast period. In contrast to 3TP, the data-driven pixel-mapping approach displays such signals with shadings of purple, indicating a comparable probability of a benign and malignant signal. Intense green pseudo-colours indicates tissue which is very likely to be healthy.

Visualisation and Classification of Lesion Masses

In figure 6.8 and figure 6.9, the pseudo-colouring and classification of temporal kinetic signals associated with voxels of the six benign lesions B1-B6 can be observed. Each block of images separated by a gray line refers to image slices of one lesion. The left 3×3 matrix of images depicts three different image slices with pseudo-colours computed from the continuous output of the LSM evaluating the *raw*-features (top row), the MSVM-L evaluating the *raw*-features (middle row) and the 3TP technique (bottom row). Pseudo-colours are only superimposed onto

voxels which belong to the lesion mass according to the manual lesion segmentation. The right 3×3 matrix of images presents the same three image slice as the left matrix, but the colours of lesion voxels illustrate the classification of the corresponding signals as malignant (red), normal (green) or benign (blue). In the 3TP based classification images, the colour reflects the signal characteristics in postcontrast period. The presence of significant wash-out is reflected by red. Green indicates an indistinct course of the signal in the late postcontrast period. The presence of a further increasing signal is reflected by blue. The same information can be observed for the six malignant cases M1-M6 in figure 6.10 and figure 6.11.

The lesion of case B1 is a fibroadenoma. The LSM and MSVM based pseudo-colourings indicate benign tissue in the lesion centre (second slice) surrounded by tissue classified as malignant but coloured with shadings of purple in the pseudo-colouring indicating an increased value for the benign component (first and third slice). This tissue characterisation concurs with the 3TP based visualisation. Regions coloured purple by the LSM and MSVM are depicted with intense green by 3TP indicating a strong signal uptake followed by indistinct wash-out characteristics. Tissue in the centre of the lesion is also characterised as benign.

The lesion of case B2 is histologically classified as a fibrous mastopathy. All three pixel-mapping methods lead to a visualisation, in which the lesion is depicted as a large mass of predominantly blue voxels indicating benign tissue. The small number of voxels coloured green by 3TP are characterised as normal by green (third slice, lower lesion margin), as benign by blue (third slice, right part of lesion) or as suspicious by purple pseudo-colours (first slice, lesion centre/second slice, left part of lesion).

The disorder marked by the radiologist in the images of case B3 is a scar for which 3TP indicates benign signals in the upper part of the lesion and indistinct or malignant signals in the lower part. The LSM characterises the kinetic signals of the lower part as malignant or suspicious (red to purple), whereas the MSVM displays voxels coloured green by 3TP with blue voxels.

The 3TP based pseudo-colouring of lesion B4 exposes mainly suspicious signals with indistinct wash-out characteristics (green). The corresponding voxels are coloured red with a weak blue component by the LSM and MSVM. Voxels presented blue in the 3TP based images are similarly characterised by both data-driven techniques.

Lesion B5 is a granuloma with signs of inflammation. The major part of benign voxels is similar characterised by all three pseudo-colouring techniques. A different assessment can be observed in the upper part of the lesion coloured green by 3TP. The LSM characterises this part with higher confidence as malignant than the MSVM.

Lesion B6 is a chronic mastitis. The 3TP based visualisation indicates malignant tissue in the lower part of the lesion. The same part is characterised as being normal (green) by the LSM and as malignant (red) by the MSVM. Voxels in the centre of the lesion, which are coloured green by 3TP, are coloured purple by the LSM and red by the MSVM.

The 3TP based colour images of the ducal carcinoma of case M1 expose a malignant centre surrounded by suspicious signals (green). The visualisation based on the data-driven pixel-mappings suggest a similar configuration of tissue sections. Differences between the LSM and MSVM based pseudo-colourings can be observed at the extension in the left part of the lesion (second image). This extension is coloured with green pseudo-colours by the LSM but with purple by the MSVM. The different assessment can also be observed in the corresponding classification images.

The major part of the scirrhous carcinoma of case M2 is consistently characterised as malignant

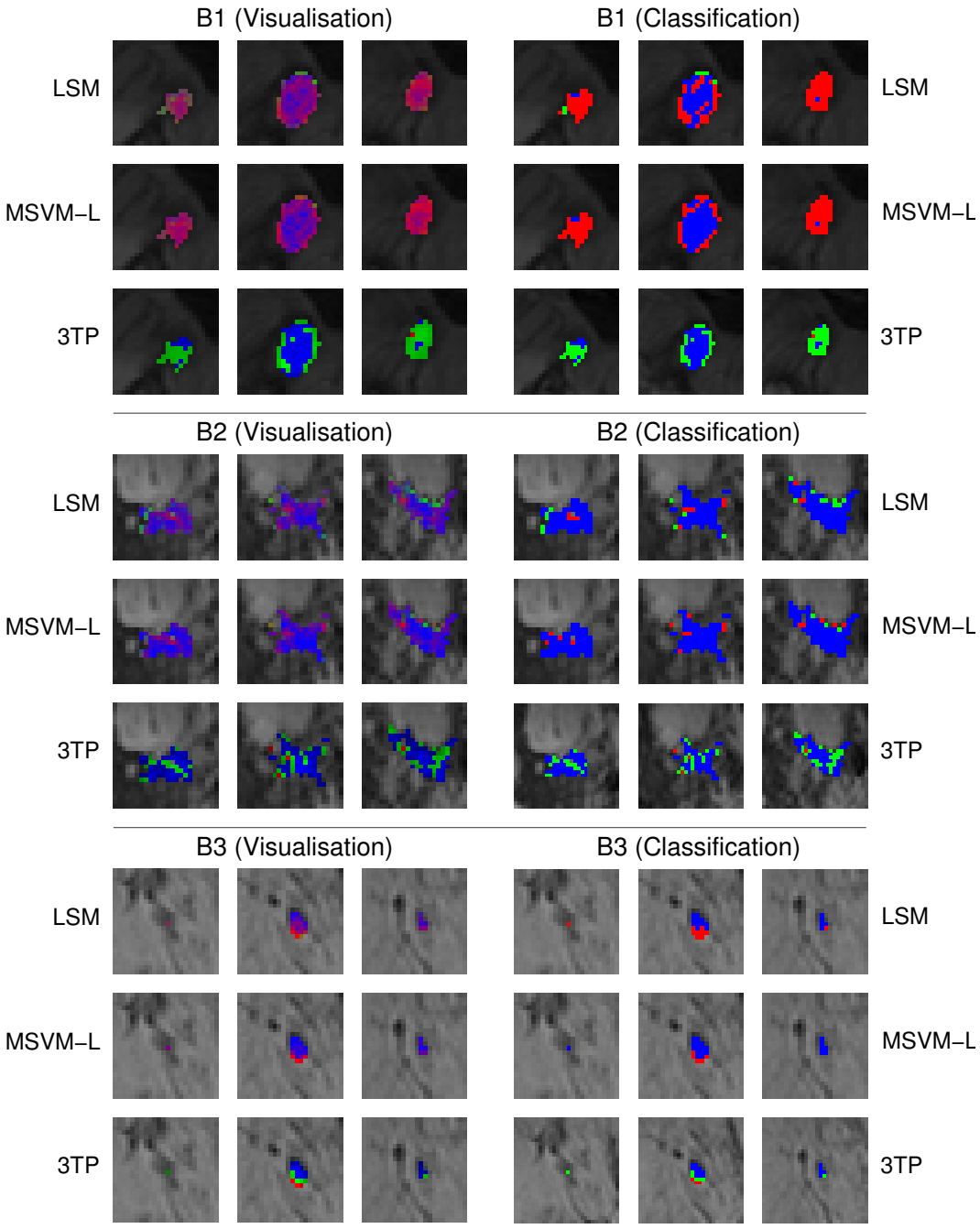


Figure 6.8: Visualisations of benign lesions based on the LSM, MSVM-L and 3TP trained with *raw*-features. Each block of images (separated by a gray line) refers to the image data depicting the lesion of the benign cases B1, B2 and B3, respectively. The left 3×3 matrix of images depicts slices with pseudo-colours computed from the continuous output of the LSM evaluating the *raw*-features (top row), the MSVM-L evaluating the *raw*-features (middle row) and the 3TP technique (bottom row). The right 3×3 matrix of images illustrates the classification of each signal as malignant (red), normal (green) or benign (blue). Pseudo-colours and classification results are only superimposed onto voxels which belong to the lesion masses according to the manual lesion segmentations.

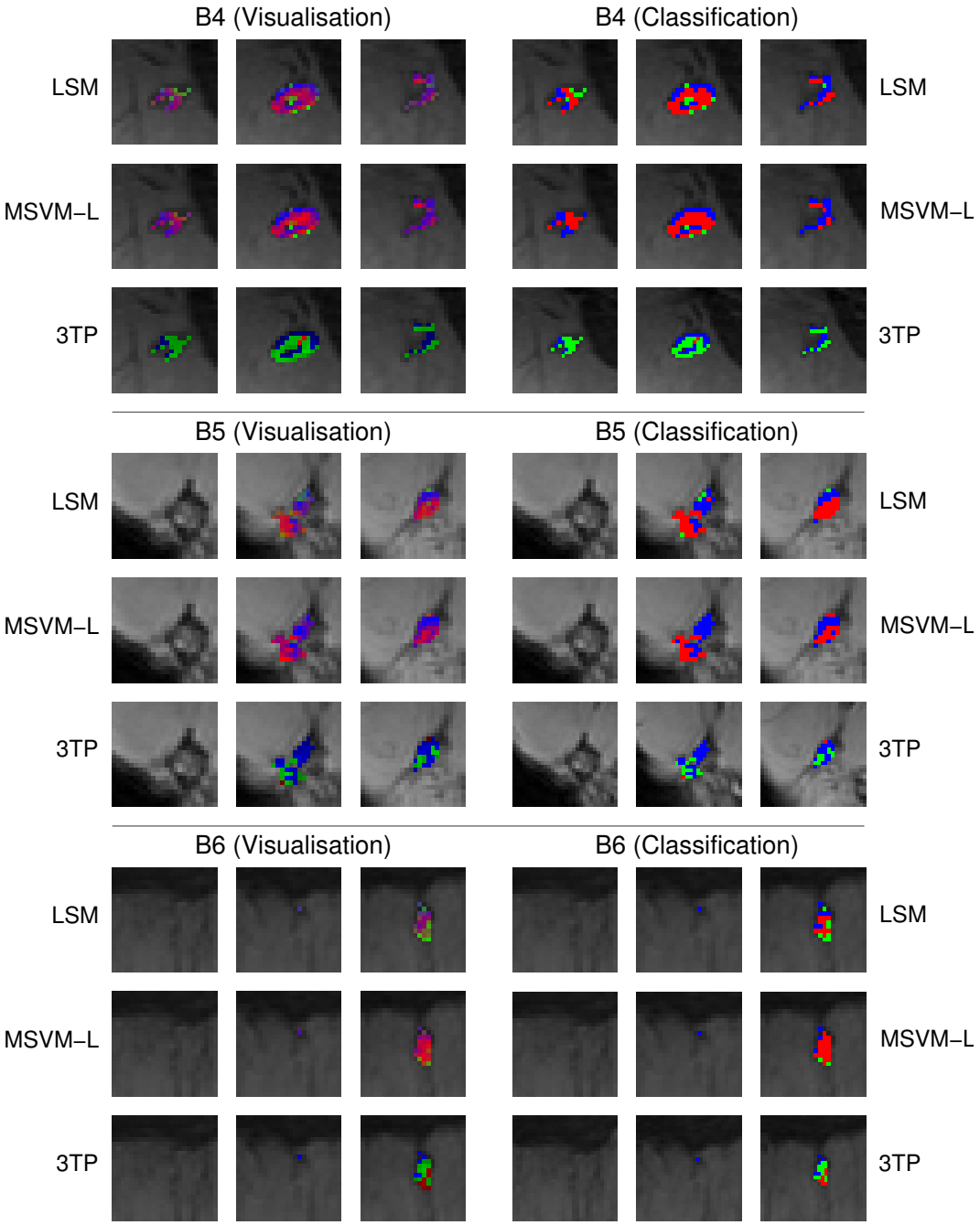


Figure 6.9: Same configuration of lesion visualisations obtained from the LSM, MSVM-L and 3TP as presented in figure 6.8 but for the three benign cases B4, B5 and B6.

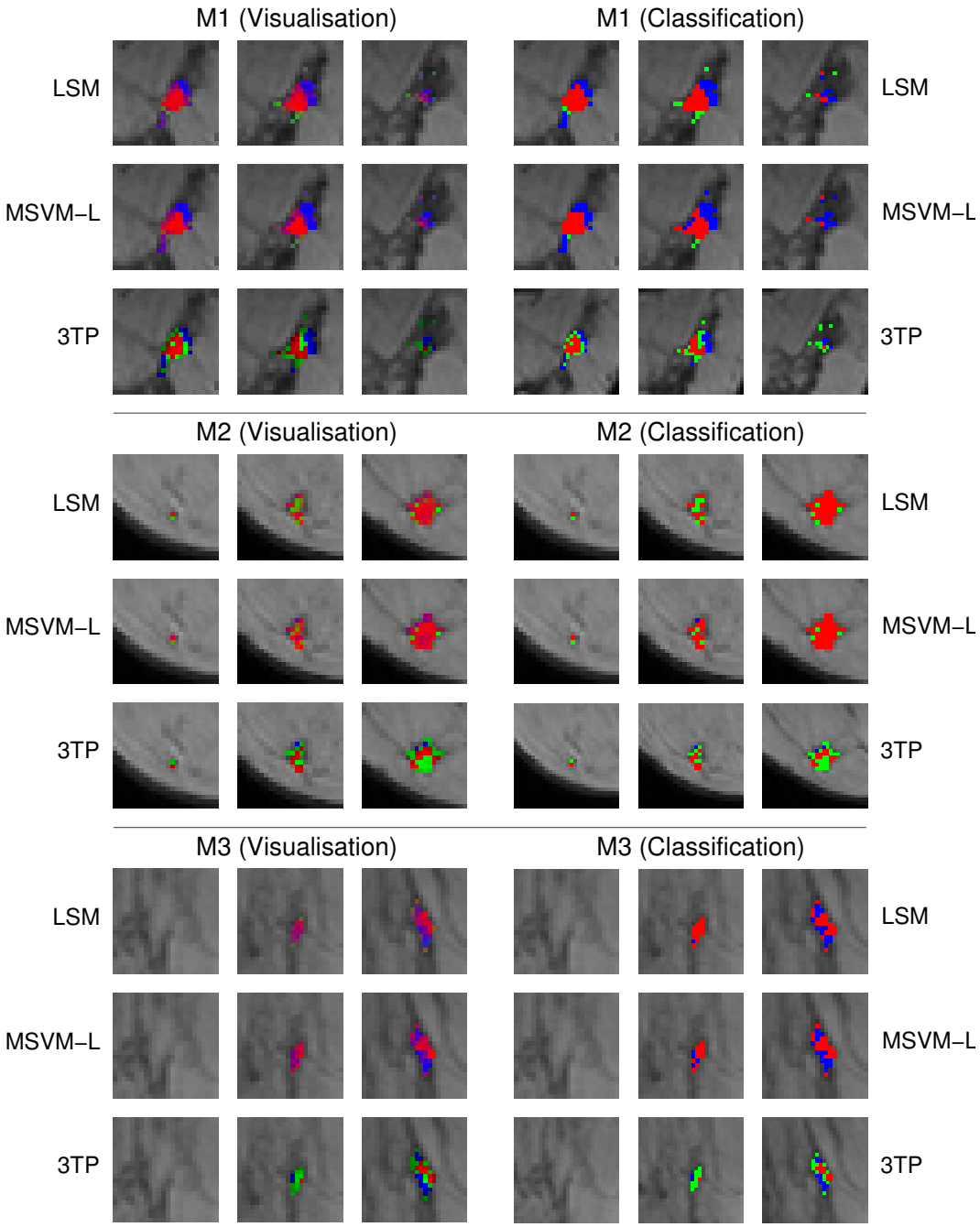


Figure 6.10: Visualisations of malignant lesions based on the LSM, MSVM-L and 3TP trained with *raw*-features. Each block of images (separated by a gray line) refers to image data depicting the lesion of the malignant cases M1, M2 and M3, respectively. The left 3×3 matrix of images depicts slices with pseudo-colours computed from the continuous output of the LSM evaluating the *raw*-features (top row), the MSVM-L evaluating the *raw*-features (middle row) and the 3TP technique (bottom row). The right 3×3 matrix of images illustrates the classification of each signal as malignant (red), normal (green) or benign (blue). Pseudo-colours and classification results are only superimposed onto voxels which belong to the lesion masses according to the manual lesion segmentations.

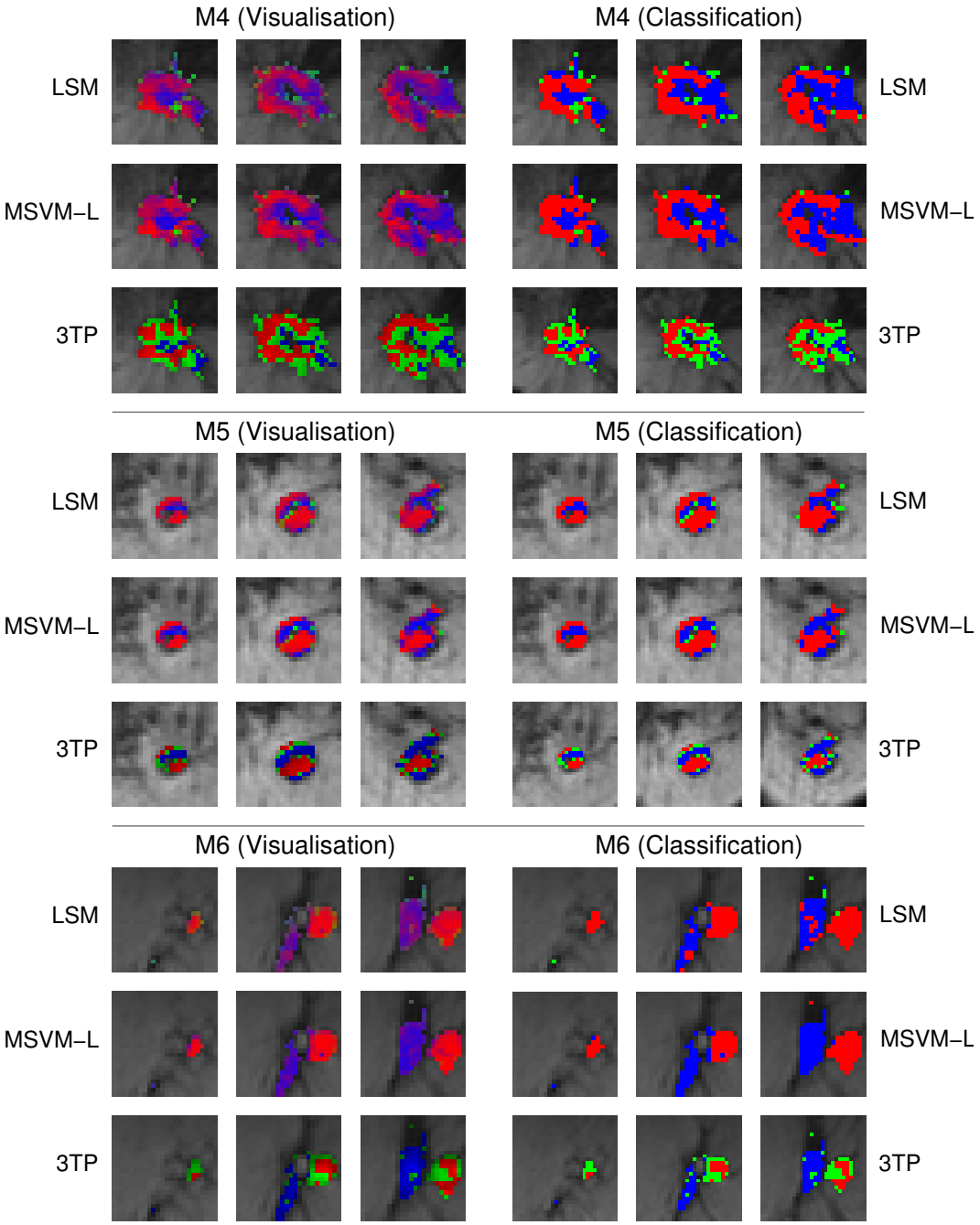


Figure 6.11: Same configuration of lesion visualisations obtained from the LSM, MSVM-L and 3TP as presented in figure 6.10 but for the three malignant cases M4, M5 and M6.

		Classifier		
		Red	Green	Blue
3TP	Red	0	2	1
	Green	0	2	0
	Blue	1	2	0

Figure 6.12: Cost matrix depicting costs associated with the different mismatches between the 3TP based classification and the outcome of the data-driven classification algorithm.

by both data-driven methods. Differences can be observed for a small number of voxels in the second slice characterised as normal (green) by the LSM but as malignant (red) by the MSVM. 3TP points to a comparable assessment of the lesion signals, even though several signals only exhibit indistinct wash-out characteristics.

The lesion of case M3 is a ductal carcinoma in situ. Both data-driven pixel-mappings suggest similar characteristics as for the preceding lesions. The upper part of the lesion, which is coloured red and green by 3TP, is coloured red to purple by the LSM and MSVM indicating a comparable characterisation of the tissue.

Lesion M4 is a multilocullar recurrent ductal carcinoma. According to the 3TP pseudo-colouring, the lesion consists of circular arranged clusters of voxels with malignant signal characteristics. The right part and the centre of the lesion exhibit only indistinct signal characteristics. The malignant part of the lesion is identically characterised by the LSM and MSVM, as well. Voxels coloured green by 3TP are coloured blue to some extent with a weak red component by both data-driven methods.

Lesion M5 is ductal papillomatosis. Both data-driven pixel-mapping techniques are in concordance with the 3TP based assessment of the lesion tissue.

The bifocal ductal carcinoma of case M6 can be divided into two parts. The right part of the lesion is characterised by 3TP as malignant with a small number of green voxels suggesting an indistinct wash-out. This part of the lesion is classified with high confidence as malignant by both data-driven techniques indicated by intense red voxels. 3TP indicates benign signal characteristics for the tissue of the left part of the lesion. Whereas the MSVM conforms with the 3TP assessment of the left part, the LSM based pseudo-colouring also suggests a small number of malignant signals.

Quantifying the Congruency Between 3TP and ANN Based Approaches

For a quantitative assessment of the data-driven evaluation of lesion masses, the classification of lesion voxels is compared to the outcome of the 3TP technique. Since it is impracticable to determine the true biological state of the tissue that caused the signal for each measured temporal kinetic signal, the quantification of the conformance with 3TP, as a clinically applied technique, enables to appraise whether the ANN based classification concurs at least with an accepted standard technique for DCE-MRI data analysis.

For deriving a label for lesion voxels from the outcome of 3TP, it is important to note that for the considered cases nearly all voxels marked by the radiologist exhibit a significant signal

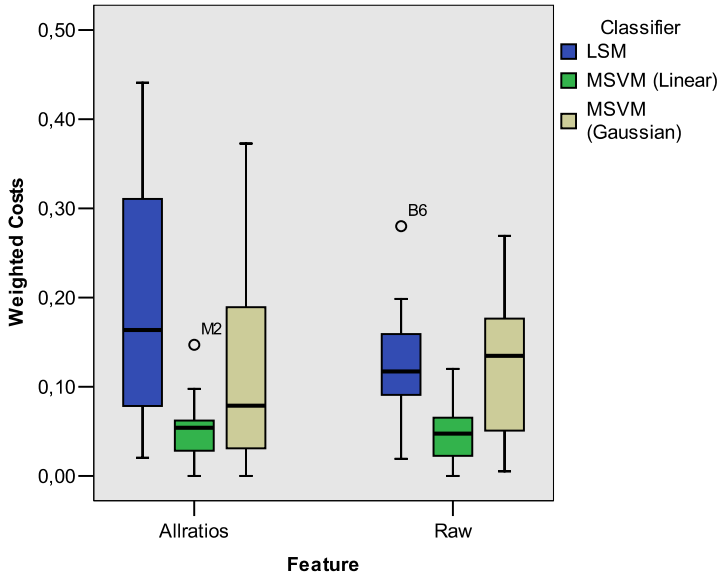


Figure 6.13: Box plot illustrating the misclassification costs caused by the different combinations of classification algorithms and feature sets. For each combination, the median, interquartile range, outliers and extreme values are presented.

enhancement during the regarded space of time. Thus, each voxel is likely to relate to tissue affected by pathological disorders. Therefore, class labels are solely assigned according to the colour hue of corresponding 3TP pseudo-colours. Voxels coloured red by 3TP indicate a significant wash-out and are likely to refer to malignant tissue. Blue voxels suggest continuously increasing concentrations of contrast agent and are likely to refer to benign tissue. Nevertheless, 3TP considers an additional class of signals with indistinct wash-out characteristics. Since the data-driven pixel-mappings provide no counterpart for this class, a customised cost matrix must be used for evaluating the confusion matrix determined for each lesion.

Since each lesion voxel exhibits a certain amount of signal enhancement, they should be classified by the ANN at least as suspicious, i.e. malignant or benign. Therefore, a 'normal' classification by the ANN is charged with 2, irrespective of the 3TP outcome for the corresponding signal. The costs are the highest in the entire cost matrix since loosing a malignant or benign part of a lesion is worse than classifying a malignant signal as benign or vice versa. The latter two types of mismatch cause costs of 1. Voxels classified as malignant or benign by the LSM or MSVM cause no costs if 3TP indicates the same class or indistinct signal characteristics (green). Table 6.12 summarises the costs associated with each type of mismatch between the outcomes of 3TP and one of the classifications algorithms.

In order to compare a certain combination of classification algorithm and feature type with the 3TP technique, a confusion matrix is calculated for each lesion. Subsequently, the costs for each lesion are determined by summing up entries of the confusion matrix multiplied with the corresponding entries of the cost matrix. The box plot in figure 6.13 illustrates costs as caused by the different combinations of classification algorithms and feature types. For each combination,

the median, interquartile ranges, outliers and extreme values are depicted. The best results are yielded by employing the MSVM-L for the evaluation of temporal kinetic signals reflected either by the *raw*-features or by the *allratio*-features.

6.4 Discussion

The visualisations presented in the preceding section underline the demand for visualisations of DCE-MRI data that take into account the heterogeneity of lesion tissue. The data-driven pixel-mapping based on supervised classification algorithms as well as the 3TP based visualisation display the lesions as heterogeneous masses of tissue exhibiting malignant, benign and normal kinetic signals. The explanatory power of the average temporal kinetic signal is obviously limited and a careful examination of the different tissue compartments by the radiologist is necessary. For this purpose, the data-driven pixel-mapping provides a visualisation of the lesion masses which facilitates the identification of tissue compartments with homogeneous signal characteristics.

The plausibility of the pseudo-colours as derived from the adaptive pixel-mapping functions was exemplarily analysed by simultaneous examination of pseudo-colours of lesion voxels and the corresponding kinetic signals (Fig. 6.7). The kinetic signal with malignant enhancement pattern was correctly depicted by an intense red voxel. The signal exposing a sustained enhancement typical for benign tissue was correctly indicated by a blue voxel. A green coloured voxel denoted a signal with an indistinct course of CA concentration. Even though the analysis of kinetic signals and the associated pseudo-colours as measured for the lesion of case M1 only exemplarily illustrates the relations between kinetic signals and the assigned pseudo-colours, the comparison of the ANN based visualisations with those obtained from 3TP reveals that both techniques widely concur in the assessment of lesion tissue. Lesion compartments which are likely to be benign or malignant are localised at similar positions. This concordance was further underlined by the evaluation of the confusion matrices reflecting the consensus between the classification of signals by 3TP and one of the ANN based pixel-mapping setups. This evaluation indicates a high concordance between both techniques, in particular if the MSVM-L is employed for evaluating unprocessed temporal kinetic signals (*raw*-features). Differences between the colouring of lesions can be observed in the presentation of those compartments which are coloured green by 3TP indicating indistinct wash-out characteristics of the corresponding signals. Such signals are predominantly coloured with shadings of purple due to similar probability values for a benign and malignant classification. A small number of such signals is also mapped to shadings of green indicating a signal time course which is more likely to be caused by normal tissue.

The 3TP based pseudo-colouring and the pseudo-colouring derived from the data-driven technique also differ in the visual contrast between pseudo-colours reflecting different signal characteristics. 3TP exclusively encolours voxels with the three primary colours red, green and blue. In particular the pseudo-colours of lesion voxels can easily be distinguished for most cases due to the substantial enhancement of the underlying tissue leading to medium to intense colours. In contrast to 3TP, the data-driven approach exploits a large spectrum of colours, and temporal kinetic signals with indistinct characteristics are reflected by mixtures of red, green and blue. Thus, it may be more difficult for the radiologist to perceive differences between certain pseudo-colours in the ANN based visualisations than in those computed with 3TP. Nevertheless, it will be

demonstrated in the following chapter that in case of the ANN based pixel-mapping, the smoothly varying characteristics of the signal time courses are consistently reflected by smoothly varying pseudo-colours, whereas the 3TP based pseudo-colouring exhibits discontinuities in the presentation of signal characteristics. For the discussion of the visual presentation of lesions, it also important to bear in mind the different type of information that is reflected by the pseudo-colours obtained from the two different techniques. Pseudo-colours derived from 3TP visually reflect the value of two explicitly defined signal features, whereas the ANN based pseudo-colours already reflect an assessment of the signals with respect to the three tissue classes.

Another factor affecting the visual contrast between pseudo-colours derived from 3TP is the normalisation of colour intensities. The intensity values of all 3TP pseudo-colours computed for the image data of one case depend on the signal exhibiting the strongest uptake, since the intensity of the pseudo-colour associated with this signal is by definition set to 255. Due to the fact that the strongest signal uptake is repeatedly determined for each DCE-MRI sequence and typically varies from case to case, the intensity of a pseudo-colour associated by 3TP with a certain signal may vary if the same signal is measured in different image volumes. Additionally, the 3TP pseudo-colouring of an entire ROI or image volume can be disturbed if the signal exhibiting the strongest uptake is afflicted by noise. In contrast to this, the pseudo-colour of a voxel derived from the ANN solely depends on the evaluation of the signal associated with this voxel and, therewith, is not afflicted by the noise disturbing temporal kinetic signals of neighbouring voxels.

According to the misclassification costs caused by employing different combinations of classifiers and features for the classification of temporal kinetic signals of lesion voxels, the MSVM-L evaluating the *raw*-features yields the best performance for the characterisation of lesion tissue if the 3TP based classification is assumed to reflect the true state of tissue. The examined visualisations of lesion masses based on the LSM and MSVM-L evince only few differences in the assessment of temporal kinetic signals. Furthermore, the comparison with 3TP indicates that both learning algorithms are able to derive implicit models of the three signal classes from a small number of cases without requiring any model assumptions about the underlying biological process. Despite the suboptimal labelling of the training data, the outcome of the adaptive pixel-mapping widely conforms with that of the model-based 3TP technique. A less uncertain label of training examples could be obtained by using e.g. a pharmacokinetical model to manually select and label temporal kinetic signals of small ROIs as described in [Lucht et al., 2002, Lucht et al., 2001]. However, in that case, the adaptive pixel-mapping would just reproduce the outcome of a second model-based technique for evaluating temporal kinetic signals. By utilising the outcome of the histological examination for the labelling of the training data, the classification algorithm derives a pixel-mapping function for evaluation of temporal kinetic signals from a technique, which itself does not consider the DCE-MRI signal for the assessment of tissue.

7 Adaptive Colour Scales for Comparison of Pixel-Mapping Techniques

In the previous chapter, two different approaches for visualisation of DCE-MRI data using pixel-mapping techniques were compared. The first pixel-mapping technique applies different types of artificial learning algorithms to compute pseudo-colours reflecting the characteristics of the courses of temporal kinetic signals. The second technique, the *three-time-points* (3TP) method, applies an explicitly defined mathematical model of the temporal kinetic signal to determine the intensity and hue of pseudo-colours. Semantically meaningful pseudo-colours can also be computed by mapping tissue parameters estimated with a pharmacokinetical model or quantitative features of temporal kinetic signals to the red, green and blue component of RGB colours. Even though all the visualisations based on the different pixel-mapping techniques are utilised for the purpose of lesion analysis, the semantic meaning of the pseudo-colours varies for different pixel-mapping methods and has to be considered for the interpretation of pseudo-colour images.

In general, a pixel-mapping can be formally defined as a function

$$M : \mathcal{S} \rightarrow \mathcal{C} : \quad \mathbf{s} \mapsto \mathbf{c}(\mathbf{s})$$

that maps a high-dimensional multivariate signal $\mathbf{s}$ from a signal space $\mathcal{S}$ to a pseudo-colour $\mathbf{c}(\mathbf{s})$, which is a point in some colour space $\mathcal{C}$. Such pixel-mapping techniques are not only applied in DCE-MRI data analysis, but also in many other fields in which multivariate image data have to be evaluated by human observers. Further examples are hyperspectral image data in the domain of medical imaging and remote-sensing or multiparameter images in the field of multiparameter microscopy. The pseudo-colours enable the observer to interpret multivariate signals in the context of their spatial appearance in the image. However, to be able to derive new information from a pseudo-colour image, it is essential that the observer has a deeper understanding for the relation between the multivariate signal and the pseudo-colour and, therewith, for the semantic meaning of pseudo-colours. Hence, before visualising any image data using a pixel-mapping function, the function itself has to be explained to the user by e.g. a suitable visualisation of M. Such visualisations are also useful for other purposes such as:

- Visual comparison of different pixel-mapping techniques e.g. for the purpose of developing new techniques or for training users interested in different pixel-mapping techniques for a specific data domain.

- Calibration of pixel-mapping techniques with one or several hyperparameters which control the mapping from the signal space $\mathcal{S}$ to the space of pseudo-colours $\mathcal{C}$ and which have to be manually adjusted by the user.

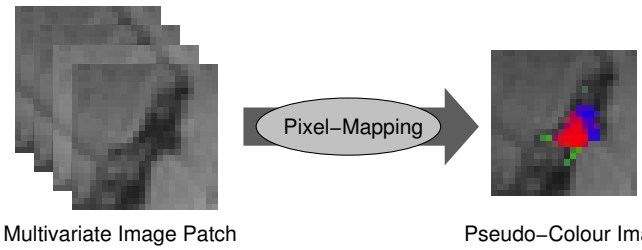


Figure 7.1: Illustration of a pixel-mapping function by means of an exemplary pseudo-colour image. The prevalent approach for introducing pixel-mapping functions is to present exemplary pseudo-colour images of small image patches displaying e.g. a lesion mass. Such images depict some of the pseudo-colours that can be obtained from a certain pixel-mapping function but expose no information about the underlying multivariate signals.

In this chapter, a method is proposed which allows for examination of pixel-mapping functions in terms of a simultaneous visualisation of multivariate signals and associated pseudo-colours. This visualisation, which is referred to as *adaptive colour scale* (ACS), is based on the application of an unsupervised artificial neural network called *self-organising map* (SOM), which has several appealing properties for this task. The ACS can not only be used in the domain multitemporal DCE-MRI data, but also in any other domain in which high-dimensional, multivariate signals are displayed as pseudo-colours.

7.1 Visualising Pixel-Mapping Functions

Visualising pixel-mapping functions is a non-trivial task due to the complex information that has to be presented. The visualisation has to depict the distribution of multivariate signals as well as the distribution of pseudo-colours in the corresponding spaces. Apparently, this is a problem of simultaneous *dimension reduction* of two different spaces, since in particular the signal space is frequently of much higher dimensionality than common two-dimensional display media. An aggravating factor is that it is essential to retain the relation between the elements of both spaces. For computing *colour-scales* reflecting the range of pseudo-colours that can be derived from a certain pixel-mapping function, one can utilise the fact that multivariate image data have a *spatial topology* and a *signal topology*. The former is given by the spatial coordinates of pixels or voxels associated with the multivariate signals, whereas the latter arises from the multivariate signal itself.

7.1.1 Colour-Scales Based on Spatial Topology

By far the most common way to illustrate a pixel-mapping function is to display pseudo-colours in the spatial topology given by the coordinates of the pixels or voxels associated with the evaluated signals. Pseudo-colours are computed for the multivariate signals and mapped back to the corresponding geometrical or anatomical sites in the image (Fig. 7.1). For instance for the data-driven pixel-mapping presented in the previous chapter, the relations between temporal kinetic signal and associated pseudo-colours were illustrated by means of a small number of exemplary images.

Figure 6.8, 6.9, 6.10 and 6.11 are examples for such visualisations. These images depict only those pseudo-colours which were computed for the temporal kinetic signals of the pixels in the displayed image patches, but neither do they provide any information about the course of the underlying temporal kinetic signals, which actually is more important for the given purpose than the spatial location of the signals, nor do they always display the entire range of pseudo-colours that can be derived from a certain pixel-mapping. Even the additional visualisation of a few signal courses as done in figure 6.7 provides only limited information about the relation between signal and pseudo-colour, because only a very small part of the variability of the signal as well as of the possible range of pseudo-colours is exposed.

7.1.2 Colour-Scales Based on Signal Topology

An alternative approach for visualising pixel-mapping functions is to compute colour-scales based on the signal topology. Pseudo-colours are not mapped to the geometrical sites of the evaluated signals in the image, but to their sites in the signal space. Thus, colour-scales based on the signal topology demand low-dimensional representations of the signal space, which are appropriate for the purpose of visualisation on a two-dimensional display.

The calibration map of 3TP (see chapter 3.5.2) can be regarded as a two-dimensional colour-scale based on the signal topology. However, the signal space and, therewith, the relation between signal and pseudo-colour is only implicitly represented by ascribing the temporal kinetic signals of DCE-MRI to two physiologically meaningful parameters of a pharmacokinetic model. By varying the two parameters of the pharmacokinetic model in reasonable ranges, artificial temporal kinetic signals are generated simulating a distribution of DCE-MRI signals. The artificial signal generated for a certain tuple of parameters is evaluated using 3TP, and the computed pseudo-colour is then depicted at the corresponding position in the two-dimensional space of model parameters. The disadvantages of this model-based approach to computation of colour-scales are that firstly, temporal kinetic signals are not displayed in their original form and, secondly, it is only applicable if signals can be described by a mathematical model, which, furthermore, is parameterised by only two input parameters.

The approach of ACSs presented in the following overcomes the mentioned shortcomings by employing the SOM algorithm for dimension reduction of the signal space. Thereby, both shortcomings of the model-based approach are circumvented: The concept of SOMs is based on learning of prototypes, which itself can be interpreted and displayed in the same way as the original multivariate signal. Additionally, the prototypes are learned from the data themselves and, therewith, do not depend on any predefined assumptions about the signal.

7.2 Adaptive Colour Scales

The fundamental concept of the ACS is the data-driven computation of a low-dimensional representation of the distribution of signals in the high-dimensional signal space. The low-dimensional form of the signal space is given by a set of *example signals* or *prototypes* $\Theta = \{\mathbf{w}_j\}, \quad \mathbf{w}_j \in \mathcal{S}, \quad j = 1, \ldots, n_\mathrm{p}$ which are arranged in a two-dimensional lattice. Each prototype is a point in the considered signal space and can be displayed, like the signals, as a line or bar plot. Additionally,

prototypes can be evaluated and associated with pseudo-colours in the same way as the recorded temporal kinetic signals.

In order to be able to analyse properties of multivariate signals by means of a low-dimensional representation, the structured set of prototypes should suffice the following requirements:

- The set of prototypes has to represent the vast majority of signal variability. It is important to note that not the entire signal space needs to be represented, but only that part filled with signals of the considered data domain. Signals of data domains such as DCE-MRI typically expose several characteristic courses from which the major part of the data exhibits only minor deviations. In this case, the data forms a manifold in the signal space and only this manifold has to be reflected by the structured set of prototypes.
- The set of prototypes has to represent the variability of signals with a certain resolution. This resolution mainly depends on the number of prototypes. The number of prototypes should suffice to represent the main types of signals such as distinct benign, malignant and normal signals in the domain of DCE-MRI data, but also those nuances of signal courses that can be observed for tissue with a less distinct assessment.
- The similarities between prototypes should be reflected by their arrangement in the lattice structure. Prototypes that are close to each other in the signal space should also be close to each other in the lattice structure. This property facilitates navigating in the visualisation of the low dimensional form of the signal space, since signal characteristics vary smoothly between neighbouring prototypes.

7.2.1 Low-Dimensional Forms of Signal Spaces Based on Self-Organising Maps

The low-dimensional representations of signal spaces are computed by applying the SOM algorithm [Ritter et al., 1992, Kohonen, 1995], an unsupervised artificial neural network which forms a topographic map of the input data. Signals which are nearby in the signal space are likely to be located nearby on this topographic map.

The two-dimensional SOM considered in this chapter consists of a lattice of n_{p} neurons. The j-th neuron of the lattice is formally described by a weight vector $\mathbf{w}_j \in \mathcal{S}$. Furthermore, each neuron is parameterised with respect to an integer coordinate pair $l_j \in \mathcal{Q}_1 \times \mathcal{Q}_2$ with $\mathcal{Q}_1 = \{1, \ldots, q_1\}, \mathcal{Q}_2 = \{1, \ldots, q_2\}$ and $n_{\mathsf{p}} = q_1 q_2$. The weight vectors can be interpreted as *prototypes of signals* located in certain subregions of the signal space and represent signal courses which are characteristical for the corresponding subsets of signals.

7.2.2 Self-Organising Maps

In order to determine the set of prototypes forming the topographic map of a given data domain, the SOM is trained with a set of unlabelled training examples $\Gamma_{\mathsf{Train}} = \{\mathbf{s}_i\}, i = 1, \ldots N$. The prototypes are first initialised with examples randomly selected from Γ_{Train}. In the online version of the algorithm, prototypes are adapted according to a random sequence of examples. For the example $\mathbf{s}(t) \in \Gamma_{\mathsf{Train}}$ selected at iteration step t, the j-th prototype is updated according to

$$\Delta \mathbf{w}_j(t) = \lambda(t) h(l_*, l_j, \sigma(t)) \Big[\mathbf{s}(t) - \mathbf{w}_j(t) \Big]$$

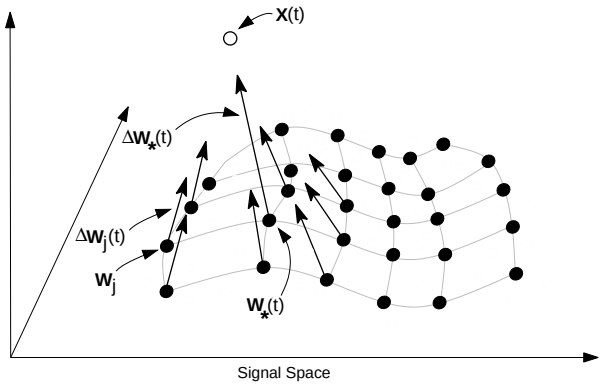


Figure 7.2: During training of the SOM, the lattice structure (reflected by gray lines) of prototypes (black discs) unfolds in the signals space according to the sequence of presented training examples. For each training example $\mathbf{x}(t) \in \Gamma_{\text{Train}}$ (circle), the best matching prototype $\mathbf{w}_*(t)$ is moved by $\Delta\mathbf{w}_*(t)$ towards the stimuli pattern. Due to the neighbourhood function, prototypes $\mathbf{w}_j$ which are topologically close to $\mathbf{w}_*(t)$ according to the lattice structure are also slightly moved towards $\mathbf{x}(t)$.

with a learning rate $1.0 \geq \lambda(t) > 0$ that decreases with the number of iterations. The *neighbourhood function* $h(l_*, l_j, \sigma(t))$ leads to stronger adaptations of those prototypes $\mathbf{w}_j$ with lattice coordinates l_j close to the lattice coordinate l_* of the best matching prototype $\mathbf{w}_*(t)$ (Fig. 7.2). The best matching prototype is determined by

$$\mathbf{w}_*(t) = \underset{\mathbf{w}_j \in \Theta}{\text{argmin}} \|\mathbf{w}_j - \mathbf{s}(t)\|$$

with a suitable metric $\|\mathbf{w}_j - \mathbf{s}(t)\|$ such as the *Euclidean distance*.

Neighbourhood Function

The neighbourhood function $h(l_*, l_j, \sigma(t))$ takes an important role in the formation of the topographical map of the signal space. The most frequently applied neighbourhood function is the *neighbourhood kernel*

$$h(l_*, l_j, \sigma(t)) = \exp\left(-\frac{\|l_* - l_j\|^2}{2\sigma(t)^2}\right)$$

with the parameter $\sigma(t)$ controlling the width of the neighbourhood. The width is chosen such that the best-matching prototype $\mathbf{w}_*(t)$ as well as the prototypes in a larger neighbourhood of $\mathbf{w}_*(t)$ are initially noticeable adapted. In the course of the training, $\sigma(t)$ is decreased leading to a decreasing weight for the adaptation of the neighbouring prototypes. For the case of $\sigma \to 0$, the update step is limited to the best matching prototype $\mathbf{w}_*(t)$ and the SOM algorithm becomes an online version of the *kmeans* algorithm. The utilisation of a neighbourhood function which initially encloses a considerable part of the prototypes and then contracts in the course of the training results in a *relaxation* or *smoothing effect* on the prototype vectors. By activating the best-matching prototype $\mathbf{w}_*(t)$ and the prototypes in its neighbourhood, i.e. those prototypes that are in the SOM lattice *topologically close to* $\mathbf{w}_*(t)$, the selected prototypes learn from the same input pattern $\mathbf{s}(t)$ leading to a *global ordering* of prototypes.

Magnification Factor

In order to obtain a low-dimensional form of $\mathcal{S}$ that allows for examination of the varying characteristics of signals, the signal distribution has to be represented by a sufficient number of prototypes. Furthermore, if signals located in certain subregions of the signal space are of particular interest, such subregions have to be exposed on the topographic map by an increased number of prototypes. This magnified view on certain subregions enables the user to examine variations of the corresponding signals at a higher level of detail.

In vector quantisation with the squared error function, it has been shown that the density of prototypes is proportional to $[p(\mathbf{s})]^{\frac{n_{\text{in}}}{n_{\text{in}}+2}}$ with n_{in} being the dimension of the signal space and $p(\mathbf{s})$ being the probability density function of the training data [Kohonen, 1995]. The inverse of this is called *magnification factor*. Even though a connection between the density of prototypes of a SOM and $p(\mathbf{s})$ has not been derived for the general case, Ritter, 1991a derived a similar power-law for the case of a SOM with a one-dimensional lattice structure. Due to this observation and the close relation to the kmeans algorithm, it can be assumed that the density of SOM prototypes roughly follows the density of the training data [Vesanto and Alhoniemi, 2000].

The conclusion that the density of prototypes roughly follows the probability density of the training data has a significant consequence for the design of the training set used for adapting the SOM. In many data domains, but most notably in domains of medical diagnosis, it can be observed that the number of normal signals n_{Normal} exceeds the number of abnormal signals n_{Abnormal} describing the phenomenon under investigation by far:

$$n_{\text{Normal}} \ll n_{\text{Abnormal}}.$$

In consequence, the prototypes of a SOM adapted with a data set reflecting the true $p(\mathbf{s})$ of the multivariate signals will predominately expose the characteristics of normal signals, whereas the phenomenon under investigation is likely to be represented by only a few prototypes. Hence, the frequency of occurrence of different signal types does not reflect their importance for the given application and need to be adjusted.

In order to increase the number of prototypes representing signals caused by the phenomenon under investigation, the probability density in the corresponding subregions of the signal space has to be increased. This can be done by e.g. replicating the corresponding signals in the training set or by reducing the number of signals that do not represent the phenomenon under investigation. Both modifications increase the statistical frequency of occurrence of the underrepresented signals in the sequence of randomly selected training examples. Therewith, prototypes are more often shifted towards the corresponding subregions of the signal space during adaptation of the SOM.

7.2.3 Visualising Adaptive Colour Scales

After adapting the SOM with a set of training signals, the lattice of prototypes forms the basis of the adaptive colour scales. Each prototype $\mathbf{w}_j$ is displayed as a chart e.g. as a line or bar plot. The background of the chart is coloured with the pseudo-colour $\mathbf{c}(\mathbf{w}_j)$ computed by evaluating $\mathbf{w}_j$ with one of the investigated pixel-mapping techniques. The n_{p} resulting charts are then arranged according to the lattice structure of the prototypes leading to a $q_1 \times q_2$ matrix of diagrams, referred to as *adaptive colour scale* (Fig. 7.3).

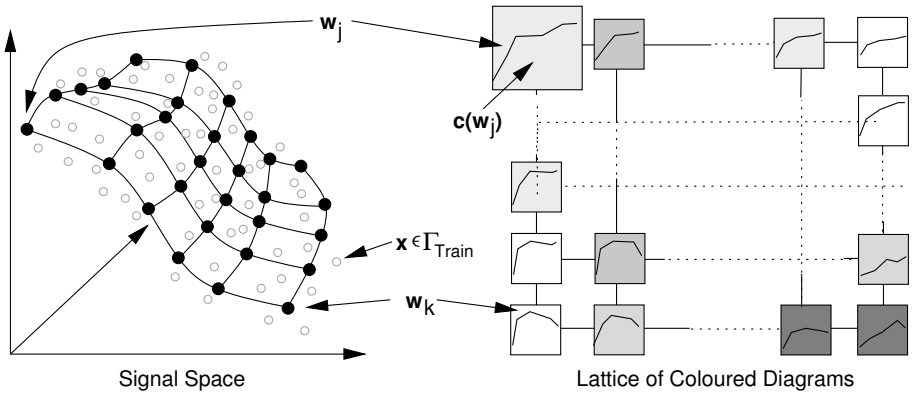


Figure 7.3: The lattice of prototypes $\mathbf{w}$, unfolded in the signal space $\mathcal{S}$ after training, is visualised as a lattice of diagrams. Furthermore, each diagram is coloured with the background colour $\mathbf{c}(\mathbf{w}_j)$ determined by evaluating the corresponding prototype $\mathbf{w}_j$ with one of the investigated pixel-mapping techniques.

7.3 Case Study: Comparison of Pixel-Mapping Techniques for DCE-MRI

In the following, an exemplary case study is presented in which ACSs are applied for comparing the data-driven pixel-mapping technique proposed in the previous chapter with the model-based 3TP pixel-mapping. To this end, a topographical map of the DCE-MRI signal space is computed by training the SOM with data of the *Munich* group of cases. Subsequently, the same SOM is successively coloured with the two pixel-mapping functions leading to the corresponding ACSs.

7.3.1 Computing a Low-Dimensional Form of the DCE-MRI Signal Space

In the domain of DCE-MRI, signals caused by pathological disorders are of particular interest but constitute only a minor part of the DCE-MRI data. In order to enhance the density of prototypes in the corresponding subregions of $\mathcal{S}$, the statistical frequency of occurrence of suspicious signals in the training set needs to be increased. To this end, the manual lesion segmentations are utilised for a selective sampling of training examples. All signals associated with lesion voxels according to the expert label of the six benign and six malignant cases of the *Munich* cohort are selected as examples for signals exposed by tissue affected by pathological disorders. The resulting set $\Gamma^{\text{Malignant}} \cup \Gamma^{\text{Benign}}$ is replicated ten-times and forms the suspicious signal fraction of the training set Γ_{Train}. Since temporal kinetic signals of normal tissue are less important in the given data domain, normal signals are added to the replicated suspicious signals with a ratio of $1:4$.

The training set Γ_{Train} is then used for training a SOM consisting of a $q_1 \times q_2$ lattice of neurons with $q_1 = q_2 = 12$. The number of neurons is chosen according to the results of preliminary experiments investigating rectangular SOMs of different size and is a trade-off between the resolution signal variations are displayed at and the space required for visualisation of the lattice structure. The SOM is adapted using the previously described online algorithm with $t_{\text{max}} = 100 \cdot |\Gamma_{\text{Train}}|$ iterations. The learning rate decreases linearly from an initial value of $\alpha(0) = 1.0$ to the final value of $\alpha(t_{\text{max}}) = 0.1$. The parameter $\sigma(t)$ controlling the width of the neighbourhood decreases linearly from an initial value of $\sigma(0) = 0.75 \cdot q_1$ to the final value of $\sigma(t_{\text{max}}) = 1.0$.

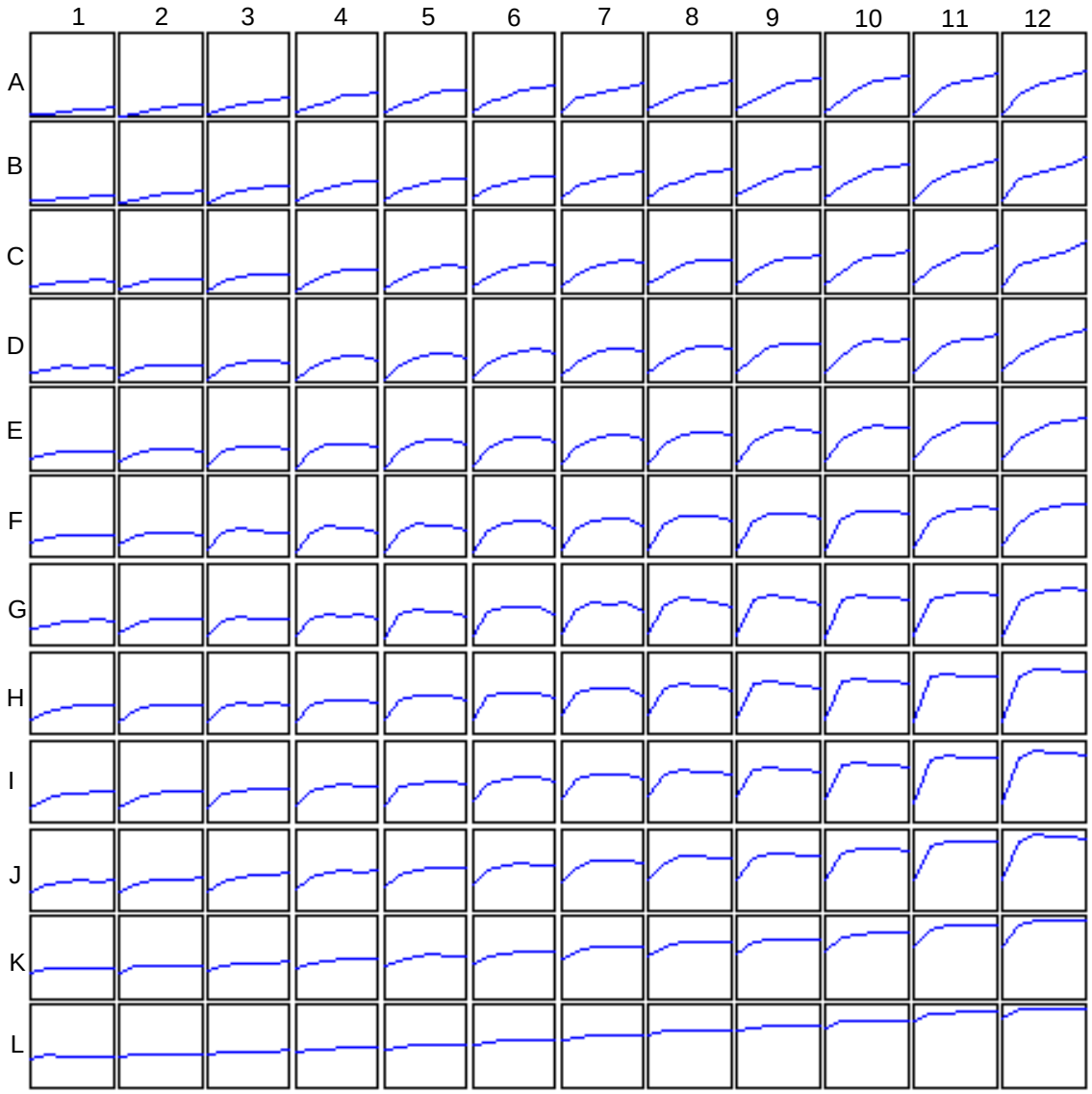


Figure 7.4: Visualisation of the 12×12 SOM trained with DCE-MRI signals. The SOM lattice is visualised as a lattice of line plots, each showing the signal course of the corresponding prototype.

7.3.2 Results

The SOM trained with the DCE-MRI data sampled from the six benign and six malignant cases of the *Munich* group can be observed in figure 7.4. Each chart of the SOM lattice illustrates the signal course of a single prototype. All three major types of temporal kinetic signals are reproduced on the SOM. In row K and L, temporal kinetic signals exhibiting nearly constant courses at different levels of intensity are depicted. Since the signals exhibit only weak signal enhancements, they are likely to be caused by different types of normal tissue. In region G-J, 6-12, temporal kinetic signals exhibiting an early signal uptake followed by a wash-out can be

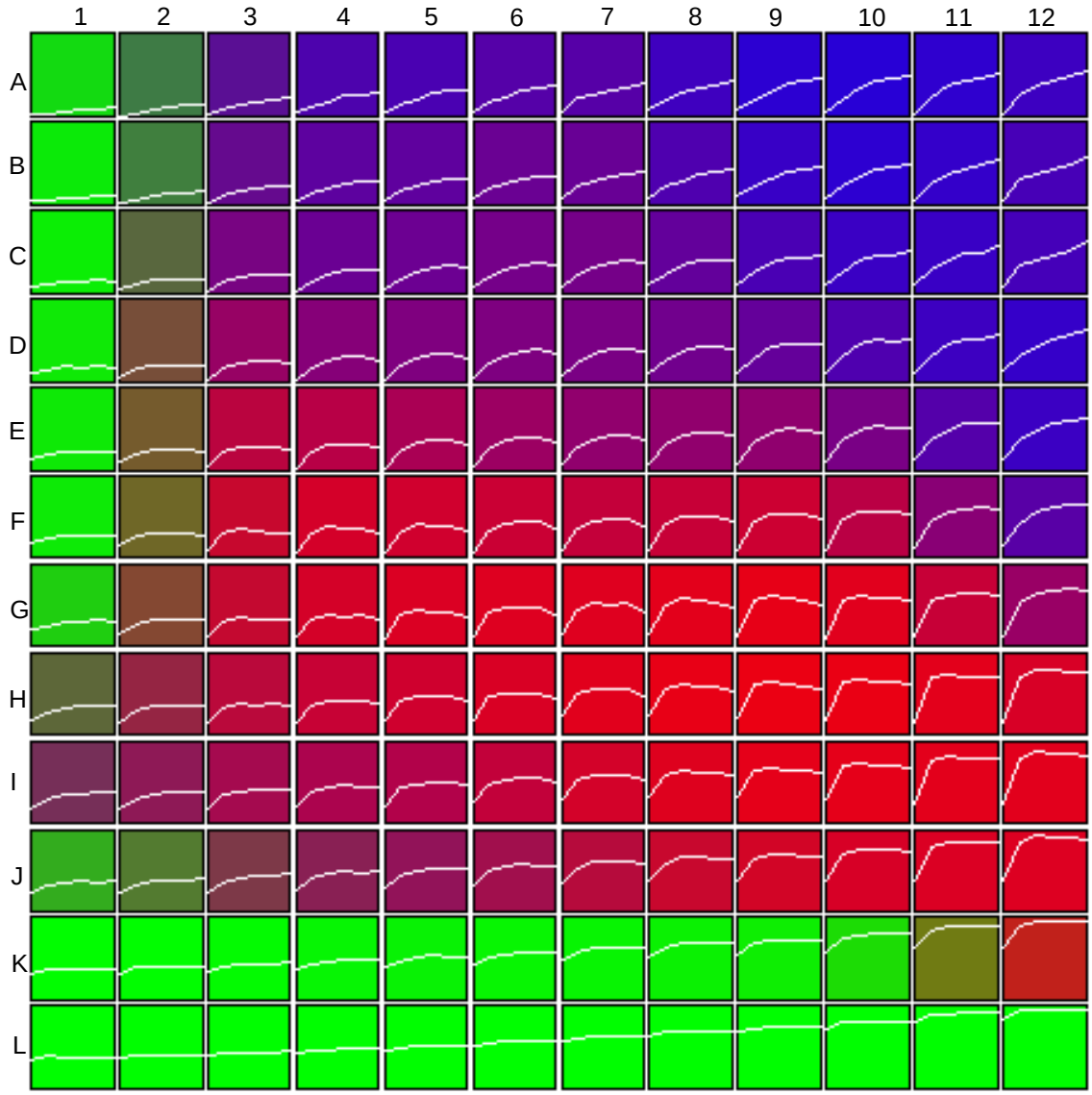


Figure 7.5: Adaptive colour scale visualising the data-driven pixel-mapping based on the MSVM with linear kernel evaluating the *raw*-features. The line plot at position x, y depicts the signal course of the prototype with lattice coordinate x, y and the corresponding pseudo-colour computed by evaluating the prototype with the data-driven pixel-mapping.

observed. Therewith, they are likely to be caused by malignant tissue. The strength of the signal uptake increases for increasing column indices. Prototypes in the region A-D, 4-12 reflect signal-time courses which are typical for benign tissue. The signals exhibit delayed signal uptakes of medium strength and signal plateaus or ongoing uptakes in the late postcontrast period.

In order to analyse the relations between the temporal kinetic signals and pseudo-colours and to compare the ANN based pixel-mapping with 3TP, the SOM presented in figure 7.4 is coloured using the two different pixel-mapping techniques. Figure 7.5 shows the ACS for the support vector

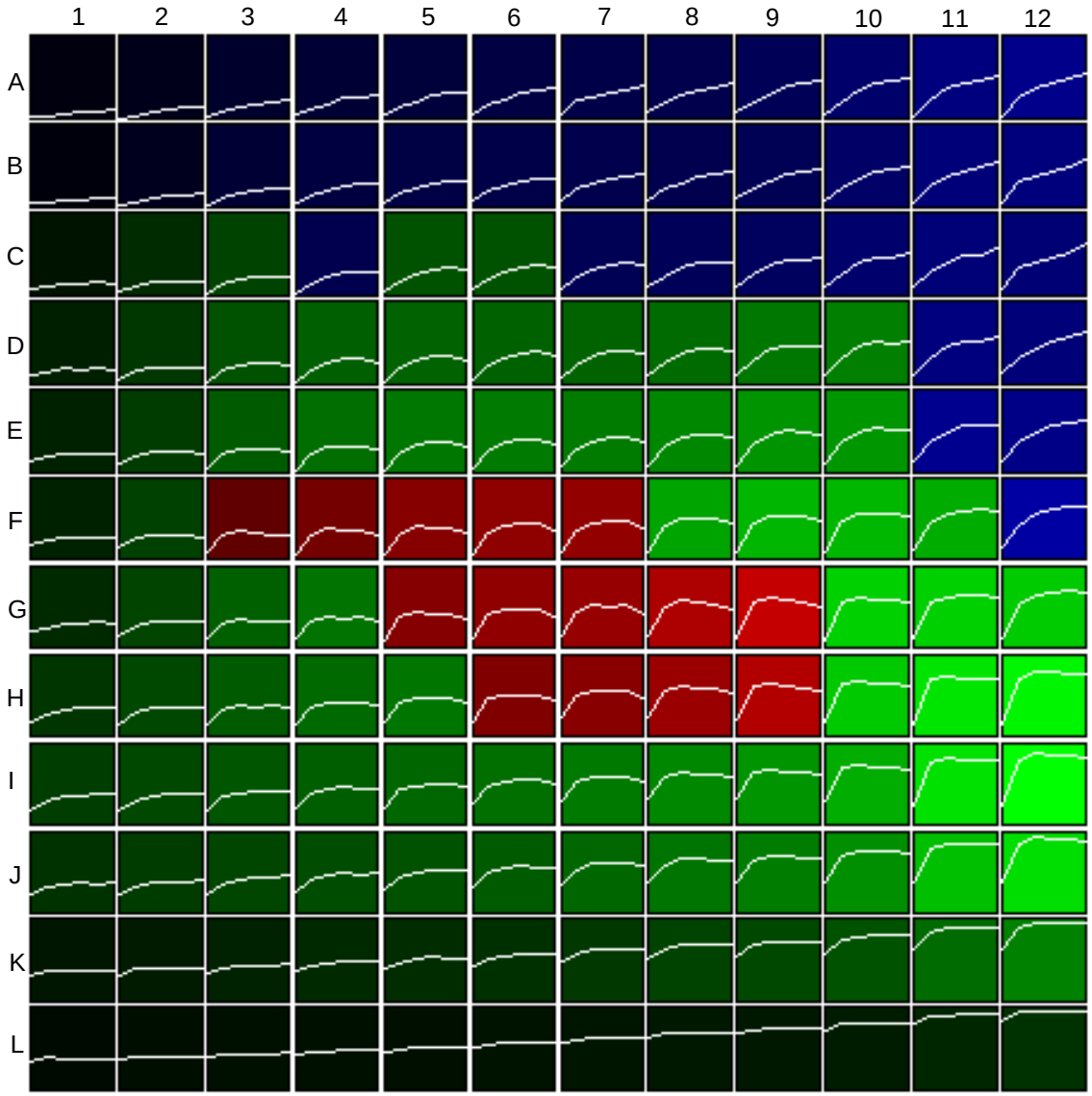


Figure 7.6: Adaptive colour scale of the DCE-MRI data coloured using the 3TP pixel-mapping. The line plot at position x, y depicts the signal course of the prototype with lattice coordinate x, y and the pseudo-colour computed with 3TP for the corresponding prototype.

machine with a linear kernel (MSVM-L) evaluating the *raw*-features (see chapter 6 for details). The ACS for the 3TP technique as it is used in the previous chapter is depicted in figure 7.6. The strength of the signal uptake and the presence of a wash-out or ongoing uptake in the late postcontrast period is evaluated by means of the precontrast and the second and fifth postcontrast image. If the signal intensity varies between the two postcontrast images more than 10%, the wash-out or continuing signal uptake is rated by 3TP as significant and displayed as red and blue, respectively. The colour intensity is scaled such that the SOM prototype exhibiting the strongest

signal uptake is displayed with an intensity of 255.

Regarding the ACS computed for the MSVM-L, temporal kinetic signals exhibiting an early uptake and a wash-out of arbitrary strength are indicated to be malignant by intense red (F-J/3-12). The more the strength of the signal uptake decreases and the more the time point of the signal peak shifts to the later postcontrast period, the more the pseudo-colour becomes purple indicating an assessment of the signal as suspicious with no clear characteristics of a benign or malignant signal (D/3-8, E/6-10). Continuously increasing signals, which are likely to be exposed by benign tissue, are displayed intense blue (upper right part of the SOM lattice).

The ACS displayed in figure 7.6 indicates that, in general, 3TP provides an assessment of prototypes similar to that of the MSVM-L. Nevertheless, only a subset of those prototypes coloured intense red by the MSVM-L is also displayed with intense red by 3TP. The major part of signals classified by the MSVM-L with high probability as malignant is depicted intense green by 3TP. The ACSs also document the concordance between both pixel-mapping techniques in the appraisement of signals which are likely to be caused by benign tissue. These signals are located by both techniques at similar positions on the ACS, i.e. in similar regions of $\mathcal{S}$. Since the strength of the signal uptake between the precontrast and the early postcontrast is typically much smaller for benign signals than for malignant signals, the major part of benign signals is depicted by 3TP with blue of medium intensity. Therewith, such signals are displayed with low visual contrast to signals coloured with shadings of dark green due to their less distinct wash-out characteristics.

The discontinuity in the 3TP based visual presentation of varying signal characteristics, which was already addressed in the previous chapter, becomes obvious in the ACS of 3TP. For certain regions of the signal space, i.e. for certain prototypes of the SOM lattice, subtle variations of the signal characteristics cause strong changes in their colour representations. Therewith, these subtle changes may also cause strong variations in the visual perception of lesions by radiologists. They may even affect the 3TP based classification of pixel clusters potentially depicting tissue disorders, since these are often classified as benign or malignant on the basis of the fractions of voxels displayed with red and blue pseudo-colours as described in [Kelcz et al., 2001]. In contrast to 3TP, the ANN based pixel-mapping colours e.g. the passage from the subregion of the signal space exposing malignant signals to the subregion of benign signals more smoothly by mixtures of red, green and blue.

7.4 Discussion

A deeper understanding for the relationships between multivariate signals and pseudo-colours is a prerequisite for analysing multivariate image data by means of pseudo-colour images. The ACS introduced in this chapter has been demonstrated to be a helpful tool for systematic examination of such relations. The visualisation of pixel-mapping functions by means of ACSs enables the observer to visually analyse the mapping of signals to pseudo-colours and, therewith, to derive knowledge about the semantic meaning of pseudo-colours in terms of signal characteristics.

Basis of the ACS is a low-dimensional representation of the signal space consisting of a lattice of prototypes. These prototypes are automatically derived from training data using the SOM algorithm. An advantage of the data-driven computation of low-dimensional forms of signal spaces is that those parts of the signal spaces that are occupied by signals are automatically

determined from the training data and predominately represented on the SOMs. A precondition for computing adequate ACSs is that the training data represent the major part of the spectrum of signals. In the case that the statistical frequency of occurrence of signals of certain classes in the training data does not reflect their importance for the given application, it is necessary to design a suitable training set by using side information such as lesion segmentations and histological reports as demonstrated in the previous section for the domain of DCE-MRI data.

In the presented case study comparing the data-driven pixel-mapping function based on ANNs with the model-based 3TP method, all three major groups of signals of the DCE-MRI domain are reflected by the SOM prototypes. Furthermore, the two-dimensional SOM also depicts less distinct signal courses combining the distinguishing features of DCE-MRI signals such as strength of signal uptake, time-point of signal peak or strength of wash-out with different nuances. The visual comparison of the two colour scales illustrates that the prototypes characterised by 3TP as malignant form a subset of the prototypes depicted as malignant by the ANN based pseudo-colouring. The ACS of the data-driven pixel-mapping further reveals that the ANN based approach also indicates malignancy for signal courses with less distinct wash-out characteristics. Furthermore, the ACSs underline the concordance between both pixel-mapping techniques in the assessment and display of signals which are benign at high probability. The two techniques mainly differ in the depiction of temporal kinetic signals with indistinct signal characteristics. These signals are coloured green by 3TP if they exhibit indistinct wash-out characteristics. The same signals are mapped to mixtures of red and blue by the ANN if they exhibit a mixture of malignant and benign signal characteristics and to mixtures of red and green if they exhibit features of normal and malignant signals. Signals coloured intense green by 3TP, indicating a highly suspicious signal uptake but unclear wash-out, are predominately displayed with red pseudo-colours in the ANN based visualisations.

In summary, the visualisations of the two pixel-mapping functions by ACSs exposes the semantic meaning of the different pseudo-colours in terms of signal characteristics, whereas the exemplary presentation of pseudo-colour images of lesions as done in the previous chapter illustrates how certain disorders are depicted by the different pixel-mapping functions in practice. Hence, both types of visualisations provide complementary information to users of such analysis techniques.

8 Image Fusion for DCE-MRI Data

Acquisition of multivariate image data, i.e. image data in which each pixel or voxel is associated with a vector of signal values, plays an increasing role in modern medical diagnostic. For instance, multispectral MRI, aggregating several three-dimensional MR images with different T1/T2-weightings, has been demonstrated to provide valuable information for e.g. diagnosing brain diseases [Holden et al., 1995,Vaidyanathan et al., 1995,Valdes-Cristerna et al., 2004]. Multitemporal images obtained from DCE-MRI allow for detecting and assessing pathological disorders of tissue in the breast, liver and prostate [Collins and Padhani, 2004]. In either case, the key information is fragmented and distributed over the entire set of images. Therefore, all images of a sequence or stack have to be analysed simultaneously rather than subsequently in order to utilise such imaging modalities to their full capacity.

Multivariate images can also be found in other fields of application. For instance in remote sensing, acquisition and analysis of multivariate images recorded with remote sensor systems such as satellites have been an area of research for several decades (see e.g. [Richards, 1993]). While the multivariate nature of most remote sensor images arises from the acquisition of multispectral data, several applications consider multitemporal image sequences for e.g. detecting temporal changes in a landscape or local events such as bush fires [Richards and Milne, 1983]. Therewith, remote sensing data bears resemblances to DCE-MRI data and the concept of *image fusion*, which has successfully been applied in various remote sensing applications, may also be helpful for the examination of DCE-MR images. While in the field of medical imaging the term *image fusion* refers to fusion of information acquired from different imaging modalities, the remote sensing community adopts a more general definition. According to Pohl and van Genderen, 1994, image fusion '*is the combination of two or more different images to form a new image by using a certain algorithm*', which also includes the combination of mutlispectral, multiparameter and multitemporal images recorded within a single imaging modality.

In DCE-MRI, a single image of a sequence provides at most morphological information about lesions and surrounding tissue, but the temporal intensity gradients, which are more important for detecting lesions, only become accessible if information from two or more images of a sequence is correlated. This observation is also the fundamental motivation for the application of fusion methods to DCE-MRI data: A suitable combination of images from a single DCE-MRI sequence may make evident signals which are not discernable in the individual images or alternatively may display the data by a reduced number of images while preserving the essential information. An important aspect is the objective of the fusion process, i.e. the question which part of the information is to be accentuated by the new images. For the field of DCE-MR image analysis, the objectives can be abstracted as:

- **Compression:** A compact visualisation of the entire DCE-MRI sequence by a single image enabling the radiologist to examine the anatomical structure of the entire breast.

- **Detection:** A visualisation of the image data accentuating spatial locations of suspicious tissue masses.
- **Characterisation:** A visualisation of the image data which enables the radiologist to differentiate benign and malignant compartments of lesions.

From the viewpoint of pattern recognition, fusion of DCE-MRI sequences is closely related to *dimensionality reduction*: Ignoring the spatial order of signal vectors induced by the associated voxels, the image data can be regarded as an unstructured set of patterns which constitute a distribution in a data or signal space. By applying dimensionality reduction techniques, the data are mapped to a new low-dimensional representation space, in which the images of the signals account for most of the structure of the original data. These low-dimensional representations of the signals can then be rearranged according to the spatial order of voxels resulting in a reduced number of images. Various artificial neural networks for dimensionality reduction have been employed for fusing image data. Villmann et al., 2003 employed the *self-organising map* algorithm [Kohonen, 1995, Kohonen, 1997, Ritter, 1991b] to fuse remote sensing data. The same algorithm was applied by Manduca, 1994 for visualising multispectral image data in a medical domain. *Principal component analysis* (PCA) [Jolliffe, 1986] was utilised by Richards and Milne, 1983 for fusing temporal sequences of remote sensing images for the purpose of bush fire detection. Another method which has been applied for the exploration of multispectral image data is the *projection pursuit* algorithm [Friedman and Tukey, 1974, Huber, 1985, Nason, 1995]. This algorithm is closely related to PCA but maximises the non-Gaussianity of the projected data, which is assumed to indicate interesting structures in the data, instead of the data's variance like in PCA as described in more detail in a latter section. A similar criterion is maximised in *independent component analysis* [Hyvärinen and Oja, 2000, Hyvärinen et al., 2001] which was recently applied for fusing DCE-MRI data [Yoo et al., 2002]. Transformation of DCE-MRI into a new set of images has also been investigated by Meyer-Bäse et al., 2004 who employed a modified ICA, the so-called *topographic ICA*.

In the following, two techniques for dimensionality reduction, PCA and the so called *kernel principal component analysis* (KPCA), are briefly described. The latter is a nonlinear extension of the standard PCA, which executes the PCA algorithm in some feature space induced by a nonlinear kernel function. Afterwards, both algorithms are employed for fusing DCE-MRI image sequences. The outcome of the fusion process is discussed with respect to the mentioned objectives *compression*, *detection* and *characterisation*. In contrast to the approaches for DCE-MRI data analysis as described in the previous chapters, application of PCA and KPCA can be regarded as *purely data-driven* approaches, since both techniques do not rely on additional label data. This can in particular be useful if a-priori knowledge about the data and the phenomenon to investigate is limited. Nevertheless, a setup in which label information is utilised to tune the KPCA based processing of image data is also investigated in this chapter.

8.1 Principal Component Analysis and Kernel Principal Component Analysis

PCA as well as KPCA can be regarded as orthogonal transformations of a basis system. Both algorithms determine a set of basis vectors, referred to as *principal axes*, which span a new representation space. The new coordinates of data examples, referred to as *principal components*, are obtained by projecting the data onto the principal axes. The difference between the two methods is that PCA is an orthogonal transformation of the basis of space $\mathcal{X}$ in which the original data is embedded[1], whereas KPCA performs an orthogonal transformation in some feature space $\mathcal{F}$ which is nonlinearly related to the original data space $\mathcal{X}$.

A convenient property of both techniques is that already a low number of principal components is often sufficient to capture the major part of structure in the data. This is due to the fact that the *intrinsic dimensionality* of data, i.e. the minimum number of free parameters of the data generating process [Fukunaga, 1990], is often lower than the dimensionality of the original signal or data space [Bishop, 1995]. Thus, the objective of PCA and KPCA is to reduce the dimensionality of the data while retaining most of the original variability in the data. Accordingly, the first principal axis lies in the direction of the highest variance and projecting data onto this axis retains the highest possible amount of variability. The principal components computed by projecting data onto each succeeding principal axes account for as much of the remaining variability as possible. It can be proven that PCA provides a sequence of best linear approximations of the data which optimally compress the signal insofar as the least-squared reconstruction error is minimised [Hastie et al., 2001]. Furthermore, principal component values computed by projecting data onto different principal axes are uncorrelated and, therefore, are likely to expose different aspects of information.

8.1.1 Principal Component Analysis

Assuming data are given by a set $\Gamma = \{\mathbf{x}_i\},\ 0 \leq i \leq N$ of examples $\mathbf{x} \in \mathcal{X}$ which are points in a n_t-dimensional data space $\mathcal{X}$. PCA performs a transformation into a new coordinate system as determined by the n_t eigenvectors $\boldsymbol{\xi}_j \in \mathcal{X}, \|\boldsymbol{\xi}_j\| = 1,\ j = 1, \ldots, n_t$ of the covariance matrix

$$C = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{x}_i - \boldsymbol{\mu}_\mathbf{x})(\mathbf{x}_i - \boldsymbol{\mu}_\mathbf{x})^T$$

with sample mean

$$\boldsymbol{\mu}_\mathbf{x} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i.$$

The eigenvectors, which correspond to the principal axes, are computed by solving the eigenvalue equation

$$\lambda \boldsymbol{\xi} = C \boldsymbol{\xi} \tag{8.1}$$

[1] In this chapter, PCA as well as KPCA are applied to unprocessed temporal kinetic signals. Thus, the signal space $\mathcal{S}$ and the data space $\mathcal{X}$ coincide. To conform with the notation of the previous chapters, examples which are designated to be processed by one of learning algorithms will be denoted as $\mathbf{x} \in \mathcal{X}$.

for eigenvalues $\lambda \geq 0$ and $\boldsymbol{\xi} \in \mathcal{X} \setminus \{0\}$ or by applying a suitable artificial neural network [Sanger, 1989, Oja, 1992, Fiori and Piazza, 1999]. By sorting the eigenvectors in the order of descending eigenvalues, an ordered basis of orthogonal and uncorrelated basis vectors can be created. The first eigenvector $\boldsymbol{\xi}_1$, i.e. the eigenvector with the largest eigenvalue, lies the direction in which the data exhibit the highest variance. The following eigenvectors are orthogonal to the preceding eigenvectors and in the direction of the highest variance in the orthogonal subspaces, respectively. Therefore, the mapping from the data space to the new representation space spanned by the first $n_{\text{rep}} \leq n_t$ eigenvectors or principal axes retains more variance than any other n_{rep} orthogonal vectors [Bishop, 1995]. The new coordinates $\tilde{\mathbf{x}}$ are computed by

$$\tilde{\mathbf{x}} = \mathbf{A}\mathbf{x}$$

with $\mathbf{A}$ being a $n_{\text{rep}} \times n_t$ matrix containing n_{rep} selected eigenvectors as rows. The components $\tilde{x}_j, j = 1, \ldots, n_{\text{rep}}$ are referred to as the *principal components* of $\mathbf{x}$.

8.1.2 Kernel Principal Component Analysis

Kernel principal component analysis [Schölkopf et al., 1998, Schölkopf et al., 1999b, Schölkopf and Smola, 2002] is a nonlinear extension of the PCA algorithm. Similar to other kernel-based methods, data are embedded in a feature space $\mathcal{F}$ nonlinearly related to the data space $\mathcal{X}$ by a nonlinear function

$$\Phi : \mathcal{X} \to \mathcal{F}, \quad \mathbf{x} \mapsto \Phi(\mathbf{x}).$$

For the time being, it is assumed that the mapped data are centred in $\mathcal{F}$. After embedding the data in the feature space, the linear PCA algorithm is executed in $\mathcal{F}$. To this end, the eigenvectors $\boldsymbol{\xi}$ of the covariance matrix

$$C = \frac{1}{N} \sum_{i=1}^{N} \Phi(\mathbf{x}_i)\Phi(\mathbf{x}_i)^T \tag{8.2}$$

have to be computed by solving the eigenvalue equation

$$\lambda \boldsymbol{\xi} = C\boldsymbol{\xi} \tag{8.3}$$

with $\boldsymbol{\xi} \in \mathcal{F} \setminus \{\mathbf{0}\}$ and $\lambda \geq 0$. Substitution of (8.2) into (8.3), resulting in

$$\lambda \boldsymbol{\xi} = \frac{1}{N} \sum_{i=1}^{N} \Phi(\mathbf{x}_i)\Phi(\mathbf{x}_i)^T \boldsymbol{\xi} = \frac{1}{N} \sum_{i=1}^{N} \langle \Phi(\mathbf{x}_i), \boldsymbol{\xi} \rangle \Phi(\mathbf{x}_i),$$

indicates that solutions $\boldsymbol{\xi}$ with $\lambda \neq 0$ lie within the span of $\Phi(\mathbf{x}_1), \ldots, \Phi(\mathbf{x}_N)$. Thus, the dual form of a solution $\boldsymbol{\xi}$ is defined as

$$\boldsymbol{\xi} = \sum_{i=1}^{N} \alpha_i \Phi(\mathbf{x}_i) \tag{8.4}$$

and we may alternatively consider the set of equations

$$\lambda \langle \Phi(\mathbf{x}_j), \boldsymbol{\xi} \rangle = \langle \Phi(\mathbf{x}_j), C\boldsymbol{\xi} \rangle, \quad j = 1, \ldots, N. \tag{8.5}$$

By combining (8.4) and (8.5) to

$$\lambda \sum_{i=1}^{N} \alpha_i \langle \Phi(\mathbf{x}_j), \Phi(\mathbf{x}_i) \rangle = \frac{1}{N} \sum_{i=1}^{N} \alpha_i \Big\langle \Phi(\mathbf{x}_j), \sum_{k=1}^{N} \Phi(\mathbf{x}_k) \langle \Phi(\mathbf{x}_k), \Phi(\mathbf{x}_i) \rangle \Big\rangle, \quad \forall j = 1, \ldots, N,$$

the eigenvalue equation becomes

$$N\lambda \mathbf{K}\boldsymbol{\alpha} = \mathbf{K}^2 \boldsymbol{\alpha}$$

with the $N \times N$ matrix $\mathbf{K}$ with entries

$$K_{ij} = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle, \quad i, j = 1, \ldots, N. \tag{8.6}$$

This form of the eigenvalue equation allows to compute the mapping to the feature space implicity by substituting a suitable *kernel function* $K(\mathbf{x}_i, \mathbf{x}_j)$ fulfilling *Mercer's theorem* [Schölkopf et al., 1999a] for the inner products $\langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle$ in (8.6) (cf. chapter 4.2.2). In this case, $\mathbf{K}$ is also referred to as *kernel matrix*.

In practice, the N eigenvectors of $\mathbf{K}$ are determined by solving

$$N\lambda \boldsymbol{\alpha} = \mathbf{K}\boldsymbol{\alpha} \tag{8.7}$$

for $\lambda > 0$. $\boldsymbol{\alpha}$ denotes the parameters of the dual form of the eigenvectors (8.4). The parameter vectors $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N$ of the eigenvectors $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_N$ can be sorted in descending order according to their eigenvalues $\lambda_1, \ldots, \lambda_N$ and provide a basis of orthogonal and uncorrelated basis vectors.

Centring Data in the Feature Space

So far, the data were assumed to have zero mean in $\mathcal{F}$. While it is difficult to centre the data in $\mathcal{F}$ explicitly, a corresponding transformation of the data set Γ can be achieved by calculating the *centred kernel matrix* [Schölkopf and Smola, 2002]. The centred kernel matrix is obtained by modifying the entries of the original kernel matrix according to

$$\tilde{K}_{ij} = (K - \mathbf{1}_N K - K\mathbf{1}_N + \mathbf{1}_N K \mathbf{1}_N)_{ij}, \quad i, j = 1, \ldots, N \tag{8.8}$$

with the matrix $\mathbf{1}_N$ with entries $(\mathbf{1}_N)_{ij} := \frac{1}{N}$. Subsequently, the centred kernel matrix $\tilde{\mathbf{K}}$ is substituted for the kernel matrix $\mathbf{K}$ in equation (8.7).

Computing Principal Components

Comparable with the linear PCA, the j-th principal component $\tilde{x}_j$ is the inner product between an example $\mathbf{x}$ and the j-th eigenvector. In contrast to the linear PCA, the inner product is not computed in the data space but in the feature space

$$\tilde{x}_j = \langle \Phi(\mathbf{x}), \boldsymbol{\xi}_j \rangle. \tag{8.9}$$

By substituting the dual form of the eigenvectors given by equation (8.4), the principal components can be expressed as weighted linear combinations of kernel functions

$$\tilde{x}_j = \frac{1}{\sqrt{\lambda_j}} \sum_{i=1}^{N} \alpha_{ji} K(\mathbf{x}_i, \mathbf{x}). \tag{8.10}$$

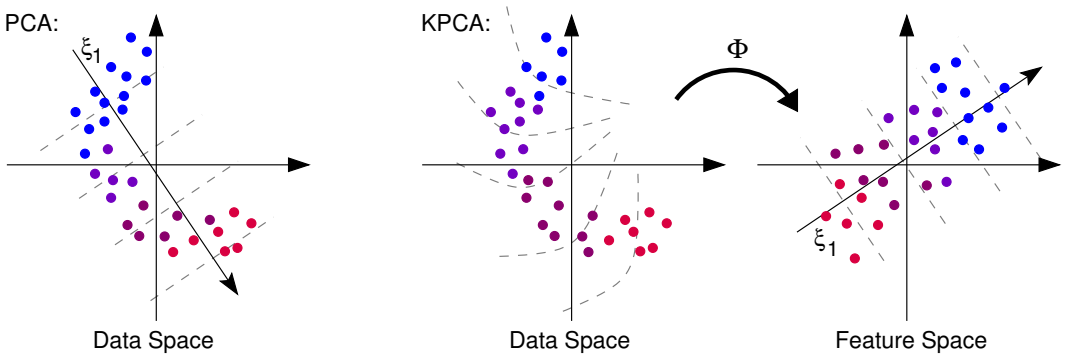


Figure 8.1: Illustration of PCA and KPCA applied to data in two dimensions. The eigenvector $\boldsymbol{\xi}_1$ determined by PCA lies in the direction of the highest data variance in $\mathcal{X}$. For PCA, projections onto $\boldsymbol{\xi}_1$ (as depicted by the dashed iso-contour lines and the colour of the data points) vary linearly in the direction of the eigenvector. If $\boldsymbol{\xi}_1$ is computed by KPCA with a nonlinear kernel function, constant projections onto $\boldsymbol{\xi}_1$ describe linear contour lines in the feature space $\mathcal{F}$ but nonlinear ones in $\mathcal{X}$ due to the nonlinear relation between $\mathcal{F}$ and $\mathcal{X}$.

The coefficient $\frac{1}{\sqrt{\lambda_j}}$ normalises $\boldsymbol{\alpha}_j$ such that the eigenvector $\boldsymbol{\xi}_j$ has unit length in $\mathcal{F}$:

$$\langle \boldsymbol{\xi}_j, \boldsymbol{\xi}_j \rangle = 1, \quad j = 1, \ldots, N$$

Choosing a Kernel Function and Its Parameters

An essential component of KPCA and other kernel-based algorithms is the transformation Φ which maps examples from the data space $\mathcal{X}$ to the feature space $\mathcal{F}$. This transformation is typically implicitly defined by the choice and parametrisation of the kernel function $K(\mathbf{x}_i, \mathbf{x}_j)$. In fact, the transformation Φ, which is implicitly computed by the chosen kernel function, even needs not to be known explicitly as long as the kernel function fulfils Mercer's theorem and, therewith, computes the inner product in some suitable feature space. If the *Gaussian kernel*

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\Big(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\Big),$$

is employed (as in this work), the user is required to set the kernel bandwidth σ to a reasonable value. Typically, the most suitable value varies for different applications. Since there is no general guideline for adjusting the kernel bandwidth in KPCA, different values are commonly tested by e.g. manually evaluating the outcome of KPCA.

8.1.3 Connection Between PCA and KPCA

According to the previous description, KPCA corresponds to executing a standard linear PCA in some feature space $\mathcal{F}$ induced by a kernel function. While the principal components computed with PCA are weighted linear combinations of the components of the original signal, the principal components obtained from KPCA are weighted linear combination of variables which are nonlinearly related to the signal components. These variables can e.g. be high-order correlations

between signal components as in the case of the *polynomial kernel of degree d* (see [Cristianini and Shawe-Taylor, 2000] for details):

$$K(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle^d.$$

Due to the nonlinear relation between $\mathcal{F}$ and $\mathcal{X}$, the constant projections onto an eigenvector in $\mathcal{F}$ describe nonlinear isocontour lines in $\mathcal{X}$ (Fig. 8.1). Furthermore, standard PCA can be regarded as a special case of KPCA if a polynomial kernel of degree $d = 1$ is used. In this case, the feature space $\mathcal{F}$ and the data space $\mathcal{X}$ coincide and the kernel function computes the inner product $\langle \mathbf{x}_i, \mathbf{x}_j \rangle$ in $\mathcal{X}$.

8.2 Fusion of DCE-MR Image Sequences

The application of PCA and KPCA for fusing DCE-MR image sequences is investigated by means of data from six patients which take part in the *MARIBS* breast cancer screening study (see chapter 3.2.1). All six cases have been diagnosed to exhibit benign or malignant tissue masses. This tissue masses were also manually segmented by a radiologist, who marked voxels associated with suspicious temporal kinetic signals with a cursor on a screening device.

For fusing DCE-MRI data, two different setups are investigated. The first setup considers the computation of *case-specific representation spaces*. Principal axes are recomputed each time a new DCE-MRI sequence is to be evaluated. To this end, PCA is applied to the DCE-MRI sequence to be fused subsequently. Thus, the principal axes and hence the semantic meaning of the fused images may vary for different cases due to variations in the individual signal distributions caused by varying fractions of normal, benign and malignant signals or even by the lack of any signal of pathologically affected tissue.

The second setup considers the computation of a single *domain-specific representation space*. Principal axes are computed once by analysing a data set representative for the given data domain. For the domain of DCE-MRI, principal axes are determined for data sampled from several training cases known to exhibit different types of breast lesions. Furthermore, the fusion process is tuned by utilising additional information given by the lesion segmentations. The information about which tissue masses are to be accentuated in the fused images of the training cases is used to build a representative data set and to optimise the bandwidth of the Gaussian kernel of KPCA. The principal axes determined from data of the training cases are then kept fixed and span the domain-specific representation space, in which data of new unseen DCE-MRI sequences are mapped thereafter.

8.2.1 Preprocessing

Each image sequence is preprocessed as described in detail in chapter 5. A binary mask, referred to as *breast mask*, is computed for each sequence. The mask is utilised to separate background voxels, which predominantly display air, from breast tissue voxels. The following fusion process is then limited to signals of breast tissue. Furthermore, signal values of each image sequence are scaled to the range $[0; 1]$. In order to avoid artefacts caused by single outlier values, the minimum and maximum intensity values of each sequence are computed after filtering each three-dimensional

image volume of the sequence with a $3 \times 3 \times 3$ median filter. The determined minimum and maximum values are then utilised for scaling the intensity values of the unfiltered image sequence. Signal values of the unfiltered images exceeding the maximum or falling below the minimum intensity of the filtered image sequence are truncated.

8.2.2 Setup I - Case-Specific Representation Spaces

In the first setup, PCA is performed on a data set Γ sampled from the current patient under investigation. This data set contains all signals marked by the breast mask (approximately $5 \cdot 10^5$ to $7 \cdot 10^5$ signals per case). The computed principal axes are then employed for mapping the same data into the new case-specific representation space.

From a theoretical point of view, it is also possible to employ KPCA in this setup. In practice, the memory requirements for storing the $N \times N$ kernel matrix $\mathbf{K}$ scale with the number of examples in Γ quadratically and computing the eigenvectors of $\mathbf{K}$, which is of $\mathcal{O}(N^3)$ complexity, becomes intractable for large data sets. While the size of Γ can be reduced by randomly selecting a subset of reasonable size, the occurrence of a large number of lesion signals in such a set is unlikely owing to the imbalance between the number of signals of healthy tissue and the number of signals of lesion tissue.

A second aspect which makes the application of KPCA in this setup unpracticable is that the outcome of KPCA strongly depends on the type and parameterisation of the kernel. In this chapter, only the Gaussian kernel parameterised by the bandwidth parameter σ is considered for fusing DCE-MRI data. Testing different values of σ for each new case is not a viable option due to the expenditure of time for the repeated computation of principal axes and components and for the visual evaluation of the fusion outcome.

8.2.3 Setup II - Domain-Specific Representation Spaces

In the second setup, PCA and KPCA are applied to a data set Γ sampled from a group of cases which are known to exhibit pathological tissue disorders and which have been evaluated by a radiologist who marked the corresponding tissue masses. Afterwards, the computed principal axes are employed for fusing image sequences of cases which have not been exposed to the PCA/KPCA algorithm before, i.e. which were not in the group of training cases.

While PCA is applied to the entire data of the training cases, KPCA is adapted and tuned using a set Γ of reduced size. For this purpose, the lesion segmentations are utilised in two ways: First, a data set Γ_{Train} is built containing a balanced number of suspicious (malignant or benign) and normal signals. The latter are randomly selected from regions marked by the breast masks excluding lesion voxels. Therewith, it is assured that Γ_{Train} always contains a considerable number of all signal types and that KPCA of Γ_{Train} remains computational manageable. Second, the knowledge about which signals are to be accentuated in the fused images of the training cases to suffice the objectives *detection* and *characterisation* is utilised for tuning the bandwidth of the Gaussian kernel.

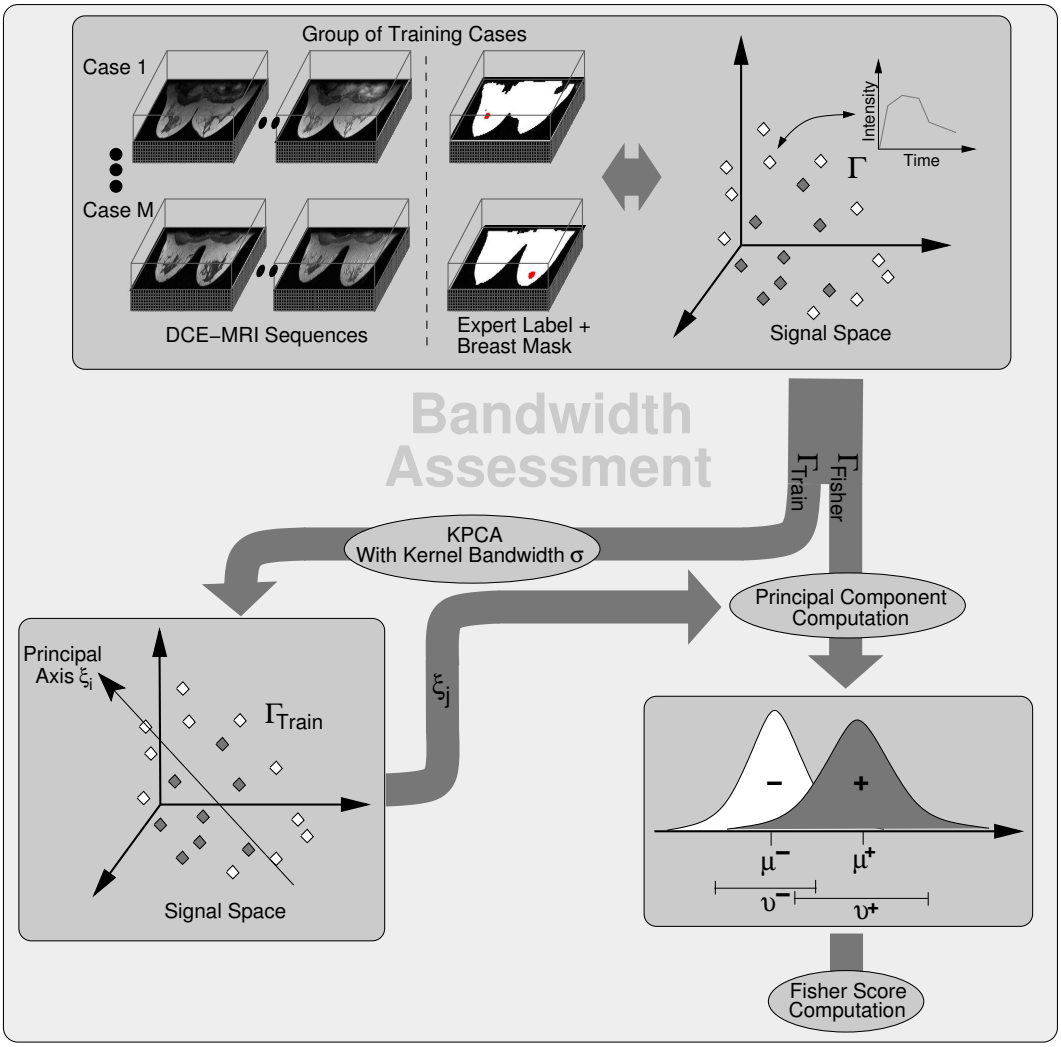


Figure 8.2: Sketch of the setup for evaluating the adequacy of a certain kernel bandwidth parameter σ for fusing DCE-MRI data. The suspicious and normal signals sampled from a group of evaluated training cases are utilised to build two sets Γ_{Train} and Γ_{Fisher} of temporal kinetic signals. Both data sets contain a balanced number of normal and suspicious signals. Principal axes are computed by analysing Γ_{Train} by KPCA with Gaussian kernel parameterised by the kernel bandwidth σ. Subsequently, the j-th principal axis $\boldsymbol{\xi}_j$ is utilised for projecting examples $\mathbf{x}_i \in \Gamma_{\text{Fisher}}$ into a one-dimensional space. The resulting principal components are subdivided according to the class label forming two class-specific distributions. The overlap between these two distributions is quantified by the Fisher's score describing the ratio of the between-class scatter and the within-class scatter. This setup is repeatedly executed for different values of σ. Those principal axes yielding the highest Fisher's score are then employed for fusing DCE-MRI sequences of unseen cases.

Selection of the Kernel Bandwidth

The bandwidth σ of the Gaussian kernel is selected by scanning the range $[0.05, \ldots, 2.0]$ using a step size of 0.05. The proceeding for evaluation of adequacy of a certain value of σ for fusing

DCE-MRI sequences is illustrated in figure 8.2. By means of a subset Γ_{Fisher} separated from Γ_{Train} before computing the principal axes, the distributions of principal component values of normal and suspicious signals are evaluated for different values of σ and different principal axes. If the overlap between the two distributions is small, it is likely that voxels of the two tissue types are displayed with high visual contrast, i.e. different gray levels, in the fused images. The overlap between the two distributions is quantified by measuring the *Fisher's score* [Bishop, 1995]

$$F_{j\sigma} = \frac{(\mu_{j\sigma}^{+} - \mu_{j\sigma}^{-})^2}{\upsilon_{j\sigma}^{+} + \upsilon_{j\sigma}^{-}}$$

with $\mu_{j\sigma}^{+}, \upsilon_{j\sigma}^{+}$ as the mean and variance of the principal components computed for suspicious signals in Γ_{Fisher} using the j-th principal axis returned by KPCA with a Gaussian kernel of bandwidth σ and $\mu_{j\sigma}^{-}, \upsilon_{j\sigma}^{+}$ as the equivalent for the normal tissue signals in Γ_{Fisher}. The overlap between the distributions is small, i.e. the Fisher's score is high, if the distance between mean values is large and, at the same time, the within-class scatter as measured as υ^{+} and υ^{+} is small. Therewith, the Fisher's score is closely related to the *Rayleigh coefficient* maximised by the *Fisher's linear discriminant* (chapter 4.3).

Beside optimisation of the kernel bandwidth, the sign of the j-th principal axis is automatically adjusted such that $\mu_j^{-} < \mu_j^{+}$ leading to fusion images in which lesion signals are more likely to be presented with higher intensity values than signals of normal tissue.

8.2.4 Displaying Fused Images

For each algorithm and each DCE-MR image sequence, the principal components based on the first three principal axes are computed and subsequently displayed as gray value images. For the purpose of visualisation, the principal component values of each image volume are scaled to the range $[0; 255]$. The gray value images computed by projecting the data onto the k-th principal axis of the standard PCA are referred to as PCA_k. Fusion images based on KPCA are referred to as KPCA_k with k denoting the principal axis onto which the data were projected. Furthermore, the three fusion images are mapped to the red, green and blue channel of the RGB composite image PCA_{RGB} and KPCA_{RGB}, respectively.

The visualisation of single fusion images can further be enhanced by mapping the range of principal component values to a suitable colour scale. Depending on the colour scale, a change of colour stimuli by the same amount may induce different changes in perception. Thus, it is advisable to employ perceptually uniform colour scales [Levkowitz and Herman, 1992, Pizer, 1981, DICOM14, 2004] for a standardised data presentation. A visualisation of a fusion image using a perceptually optimised and device-independent colour scale is shown in figure 8.3. The reader is referred to Saalbach et al., 2004 for further details about the utilised sRGB based colour scale. In order to facilitate the comparison of fusion images derived from the setups described above, only gray value images are presented in the remainder.

8.3 Results

In the following, fusion images based on case- and domain-specific representation spaces computed by PCA and KPCA are presented. The suitability of fusion images for the purpose of lesion

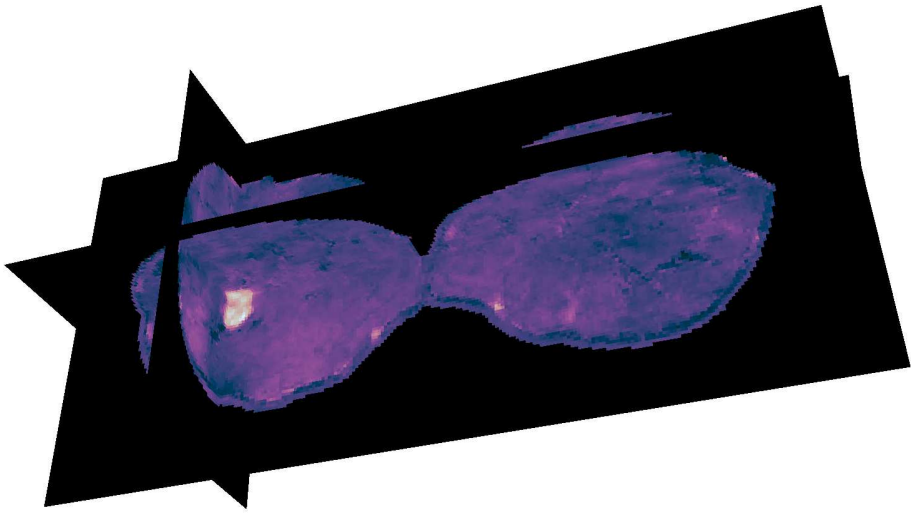


Figure 8.3: Visualisation of a fusion image using a perceptually optimised colour scale. The colour reflects the projection of $\mathbf{x}$ onto the second principal axis $\boldsymbol{\xi}_2$ computed by PCA. The lesion is located in the crossing point of the three planes parallel to the three spatial axis.

detection is quantitatively evaluated by ROC analysis. To assure that despite of the low number of cases the experimental setup reflects circumstances of a practical application, fusion of DCE-MRI sequences based on a single domain-specific representation space is evaluated following the leave-one-case-out scheme which was already used in the previous chapters. Principal axes are computed using data of five cases and subsequently employed for fusing the image sequence of the sixth unseen case. This scheme is repeated for all permutations of the six cases. The results are exemplified on the basis of the image sequences from two cases, but similar findings can be observed for the remaining cases.

8.3.1 Setup I - Case-Specific Representation Spaces

Figure 8.4 exemplarily presents the fusion images $\text{PCA}_1, \ldots, \text{PCA}_6$ computed for a single coronal slices of case M005A (left column) and M009A (right column). The principal component values of each image were scaled to the range of $[0; 255]$ for visualisation purposes. The green lines depict the contours of the two malignant lesions. The charts in figure 8.5 illustrate the corresponding eigenvalues. For both cases, more than 95% of the total data variance is retained by the first principal axis. The eigenvalues $\lambda_4, \ldots \lambda_7$ indicate that $\boldsymbol{\xi}_4, \ldots, \boldsymbol{\xi}_7$ reflect a negligible amount of variance (less than 2%). Thus, the projections into the subspaces spanned by $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_3$ already retain the major part of the total data variance of the DCE-MRI sequences. This finding can also be observed in the fusion images themselves. For both cases, the structure of the entire breast, mainly consisting of blood vessels, fat and glandular tissue, is reflected by PCA_1. The lesion masses are not accentuated in PCA_1 but can clearly be distinguished from normal tissue in PCA_2 and PCA_3. Fusion images based on the remaining principal axes provide only minor information about location and structure of the lesion mass and about the anatomical structure of the entire breast as well. PCA_4 and PCA_5 of case M005A are likely to represent long range

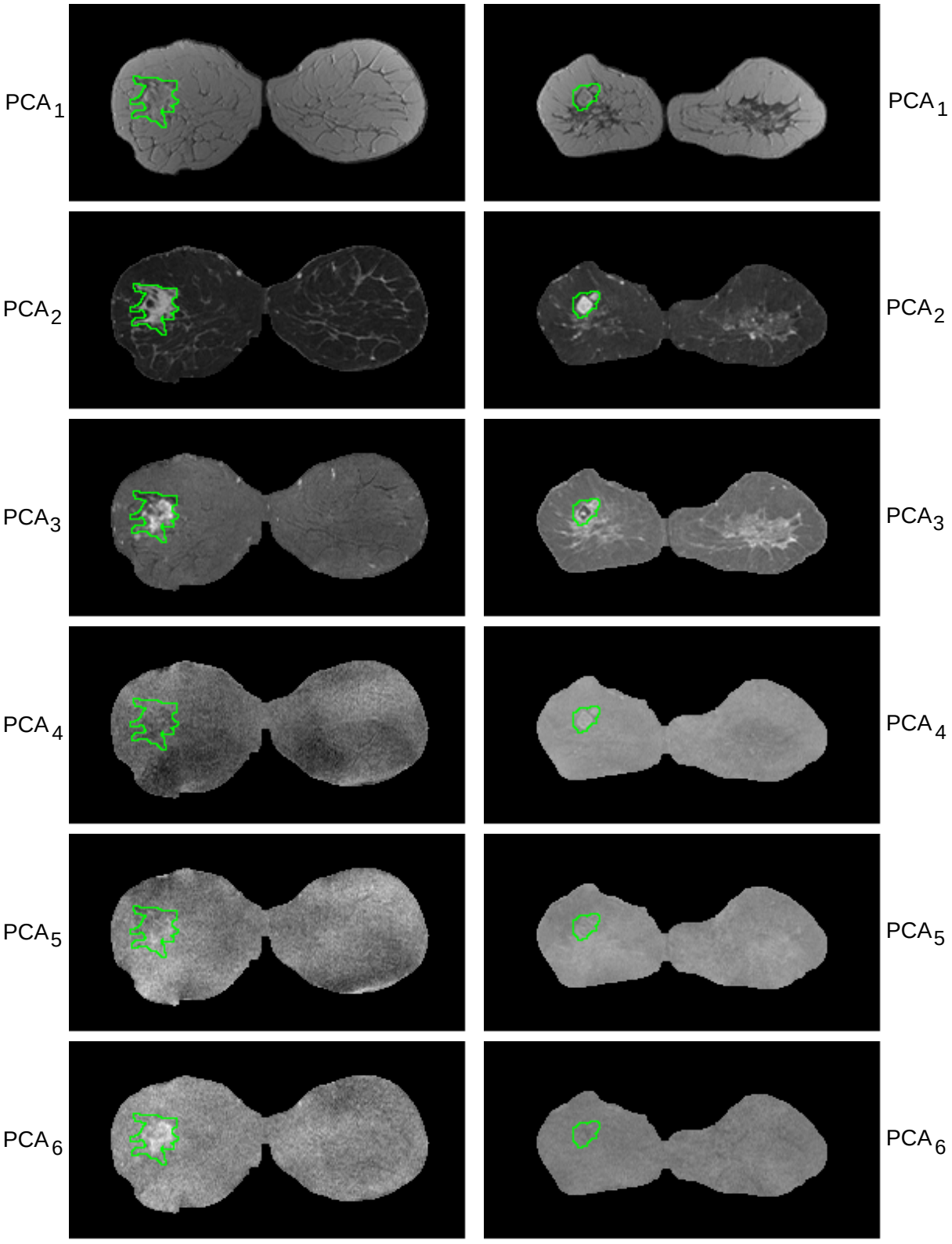


Figure 8.4: Fusion images $PCA_1, \ldots, PCA_6$ computed in setup I for cases M005A (left column) and M009A (right column). The green lines reflect the contour of the manual lesion segmentations.

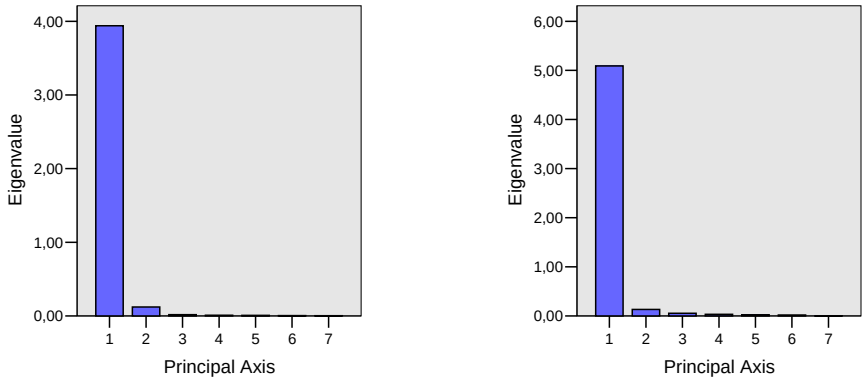


Figure 8.5: Plot of the PCA eigenvalues $\lambda_1, \ldots, \lambda_7$ computed in setup I for cases M005A (left) and M009A (right). For both cases, the major part of the data variance is captured by the first three principal axes.

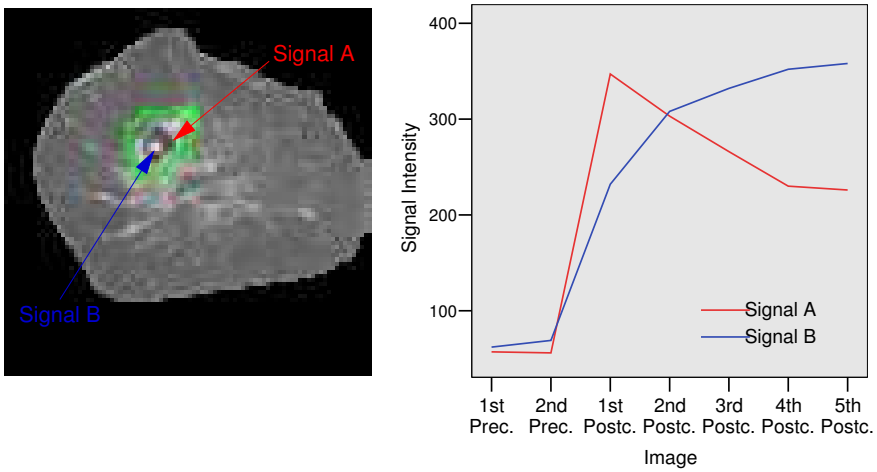


Figure 8.6: Magnified display of a subregion of PCA_3 of case M009A (left). The two arrows point to the locations of the two temporal kinetic signals presented in the right chart. Signal B as measured for a voxel of the lesion's core exhibits benign signal characteristics, whereas voxels of the surrounding ring displayed with lower intensity values expose malignant signal characteristics (Signal A).

intensity variations as caused by inhomogeneities in the magnetic field.

In PCA_2 of both cases, the lesion masses are displayed with high intensity values but do not fill the entire contour of the corresponding lesion segmentation. Nevertheless, the temporal kinetic signals associated with lesion voxels displayed with low intensity exhibit only unsuspicious signal courses, i.e. they expose no indications for pathological disorders of the underlying tissue in terms of temporal kinetic signals. This finding indicates that for segmenting lesion masses radiologists do in fact not only consider information about temporal kinetic signals but also morphological properties of larger tissue regions which are not reflected by the signals of individual voxels as evaluated by PCA.

Focusing on the visualisation of the lesion of case M009A, the fusion image PCA_3 allows to subdivide the tissue mass that is displayed homogenously with high intensity values in PCA_2. A

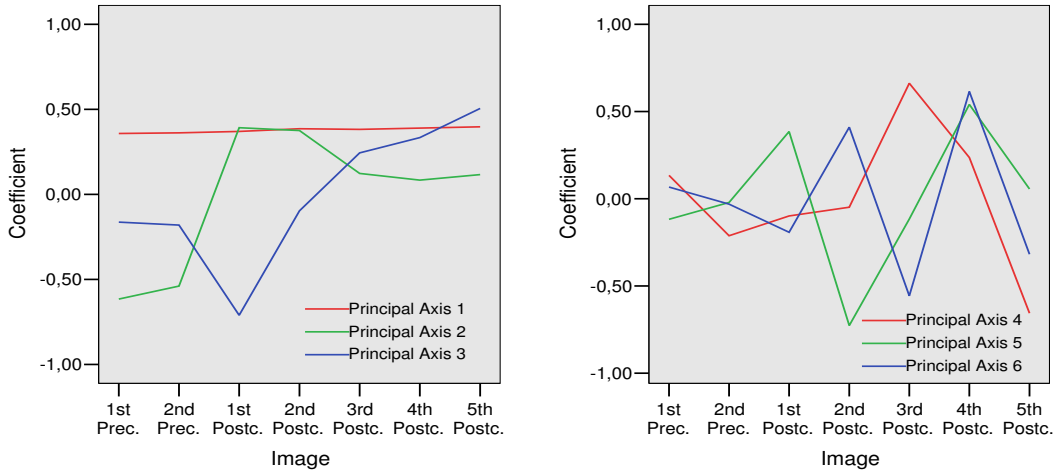


Figure 8.7: Principal axes $\xi_1, \ldots, \xi_3$ (left) and $\xi_4, \ldots, \xi_6$ (right) computed in setup I for case M009A. Axes $\xi_1, \ldots, \xi_3$ expose signal characteristics which are typical for normal, malignant and benign tissue, respectively.

magnified view of this mass is depicted in figure 8.6. The image exposes a ring pattern of low intensity values surrounding the lesion core displayed with high intensities. The chart to the right of the image illustrates the signal courses as measured at the two positions marked by arrows. The sustained increase of CA concentration as exposed by signal B indicates benign tissue within the lesion's core. The signal measured for a voxel of the ring pattern (Signal A) indicates malignant tissue due to a fast uptake of CA concentration followed by a distinct wash-out.

Figure 8.7 depicts the coefficient of the principal axes $\xi_1, \ldots, \xi_6$ computed for case M009A. Since PCA only determines the direction of eigenvectors, their signs have been manually adjusted such that lesion tissue is presented with high intensities in the corresponding fusion images. The first principal axis refers to a nearly constant signal. Thus, the major fraction of the data variance can be ascribed to temporal kinetic signals of normal tissue which constitute the major part of the image data and typically expose courses that vary only slightly at different levels of intensity. The conspicuous display of the lesion mass in PCA_2 can be ascribed to the malignant characteristics of signal course reflected by ξ_2. The signal course of ξ_3 possesses sustained signal enhancements in the postcontrast images and accounts for the accentuation of benign signals in PCA_3. The remaining principal axes cannot be related to clear semantic meanings, which is also reflected by the corresponding fusion images predominately depicting noise.

Figure 8.8 depicts the fusion images PCA_1, PCA_2 and PCA_3 of the remaining cases M007A, B015A, M094A and M104A. In general, the same image characteristics can be observed as for case M005A and M009A.

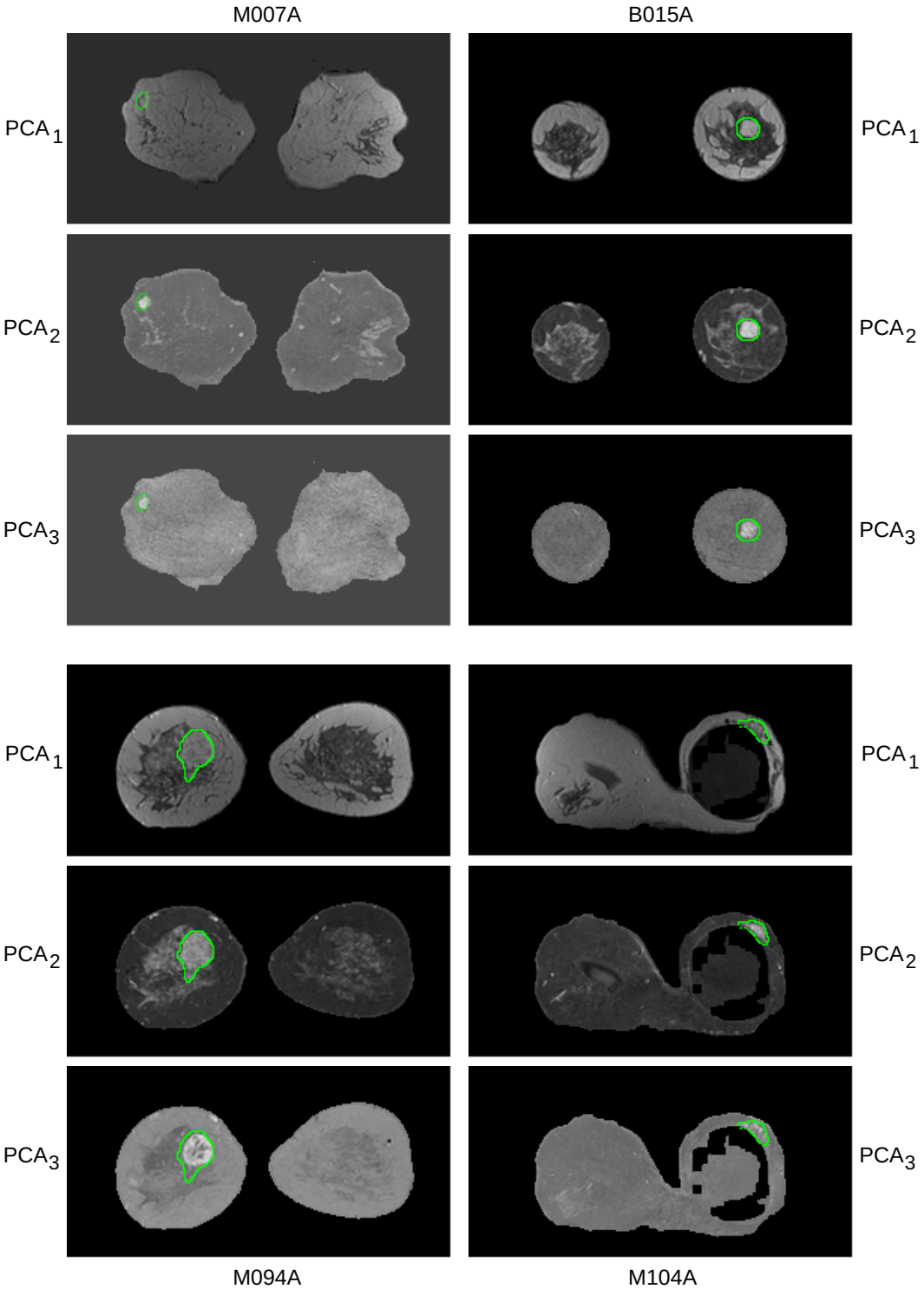


Figure 8.8: Fusion images $PCA_1, \ldots, PCA_3$ computed in setup I for case M007A (left column, first three images), B015A (right column, first three images), M094A (left column, last three images) and M104A (right column, last three images), respectively. The green lines depict the contours of the lesions as marked by the radiologist.

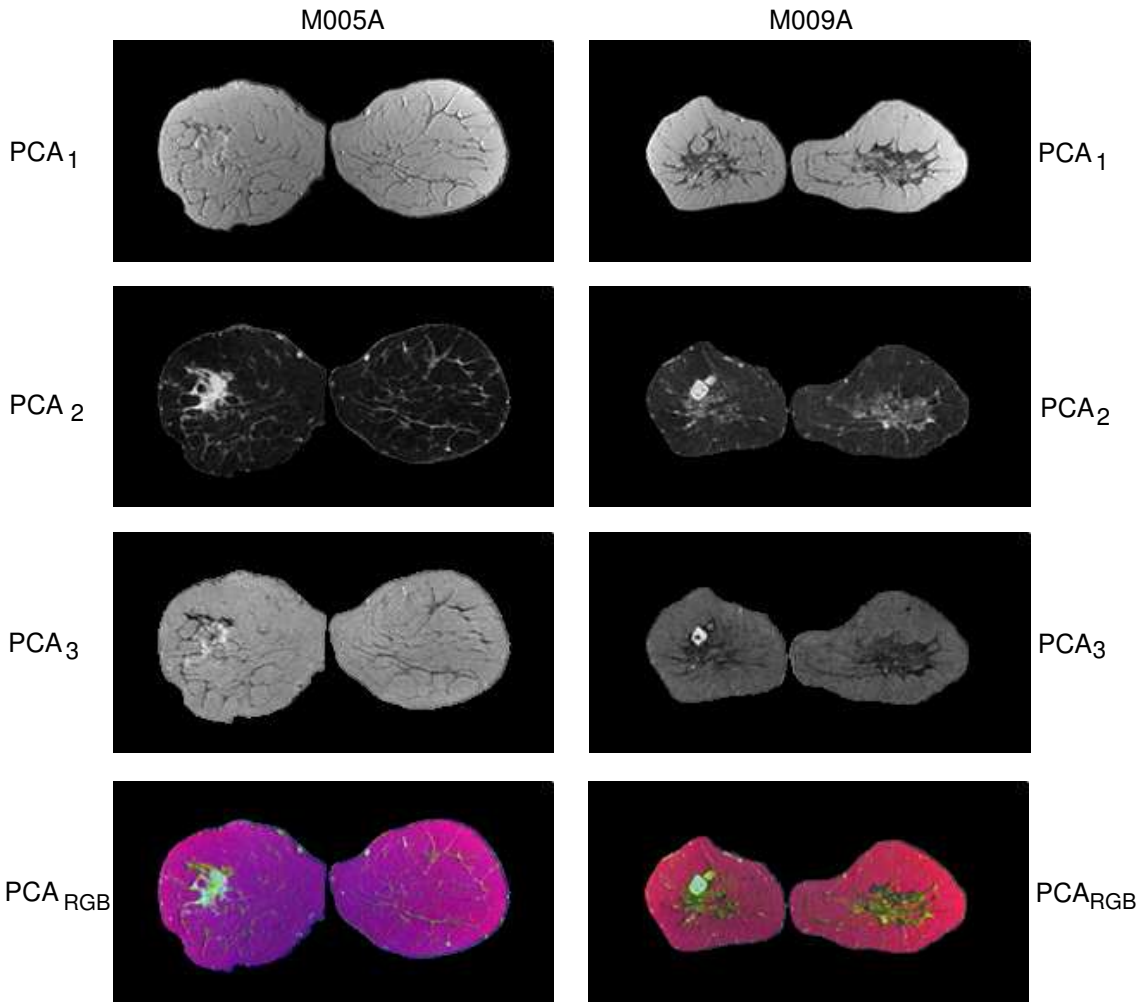


Figure 8.9: Fusion images $PCA_1, \ldots, PCA_3$ and PCA_{RGB} computed in setup II for case M005A and M009A. The composite image PCA_{RGB} combines $PCA_1, \ldots, PCA_3$ by mapping the three fusion images to the red, green and blue component, respectively.

8.3.2 Setup II - Domain-Specific Representation Spaces Based on PCA

Examples for fusion images $PCA_1, \ldots, PCA_3$ based on the domain-specific representation space determined by PCA of the given group cases excluding the subsequently fused case (M005A or M009A) are presented in figure 8.9. The presented examples are also representative for the remaining cases. Similar to the fusion images presented in figure 8.4, PCA_1 allows for inspection of the breast's anatomical structure, while PCA_2 depicts the lesion masses with high contrast to healthy tissue. PCA_3 of case M009A likewise allows to distinguish the small benign tissue compartment of the lesion from the remaining lesion tissue exhibiting malignant signal characteristics.

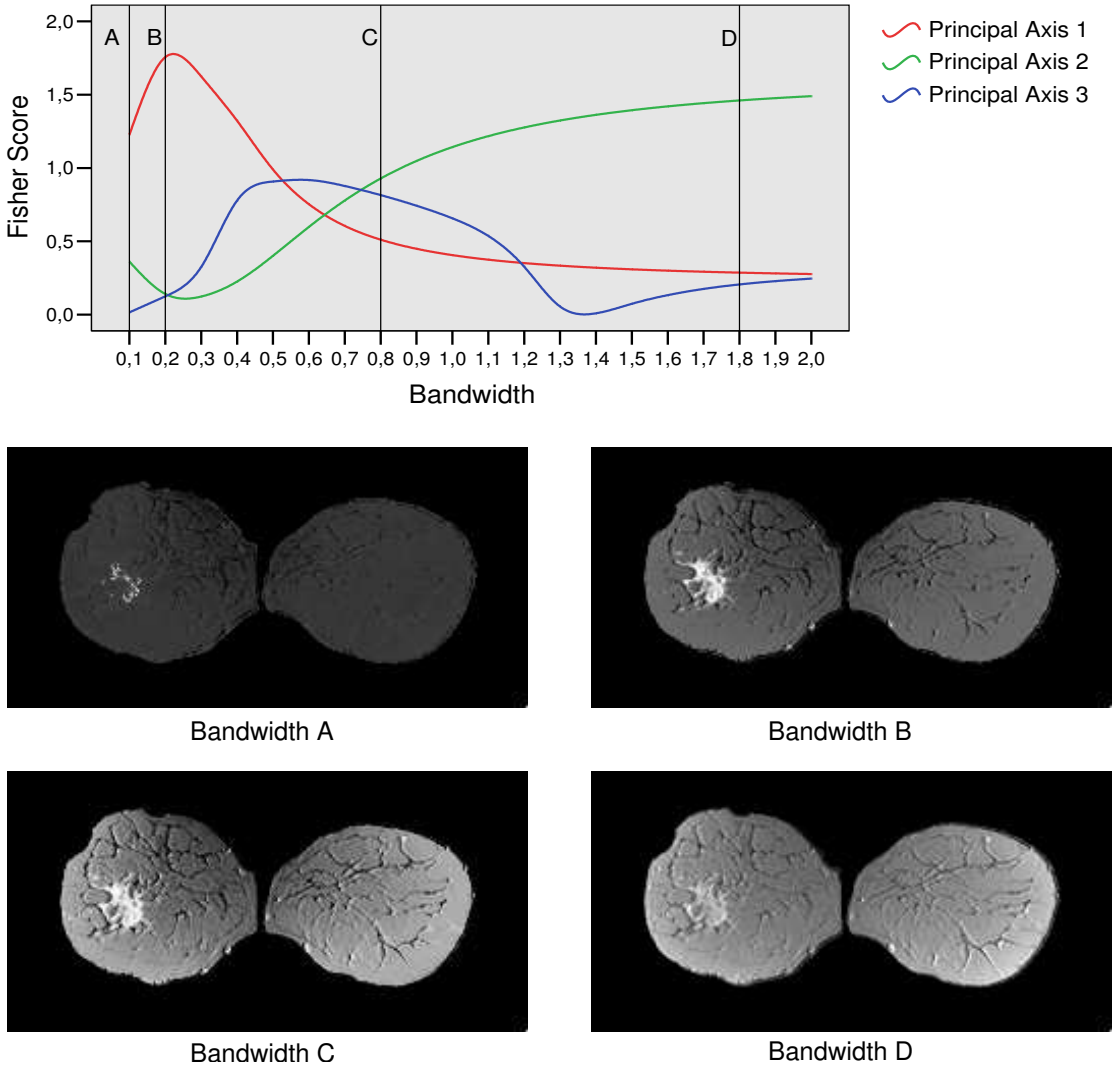


Figure 8.10: Fisher's score and fusion image KPCA_1 for different bandwidth values. The Fisher's score is determined for the principal components computed for a subset of the training data using the first three principal axes obtained from KPCA. The four images show the outcome of a fusion process, in which the first principal axis determined by KPCA with four different bandwidth values A, B, C and D is applied to fuse the image sequence of an unseen case (M005A). The highest visual contrast, i.e. the largest difference between intensity levels of lesion and non-lesion voxels, is obtained for bandwidth value B which also yields the highest Fisher's score.

8.3.3 Setup II - Domain-Specific Representation Spaces Based on KPCA

In figure 8.10, the Fisher's score is plotted as a function of the kernel bandwidth σ. The Fisher's score was evaluated for principal component values computed for examples in Γ_{Fisher} using the principal axes $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_3$ derived from KPCA of the data from the training cases excluding M005A. The images below the plot depict the fusion image KPCA_1 of the excluded case M005A computed

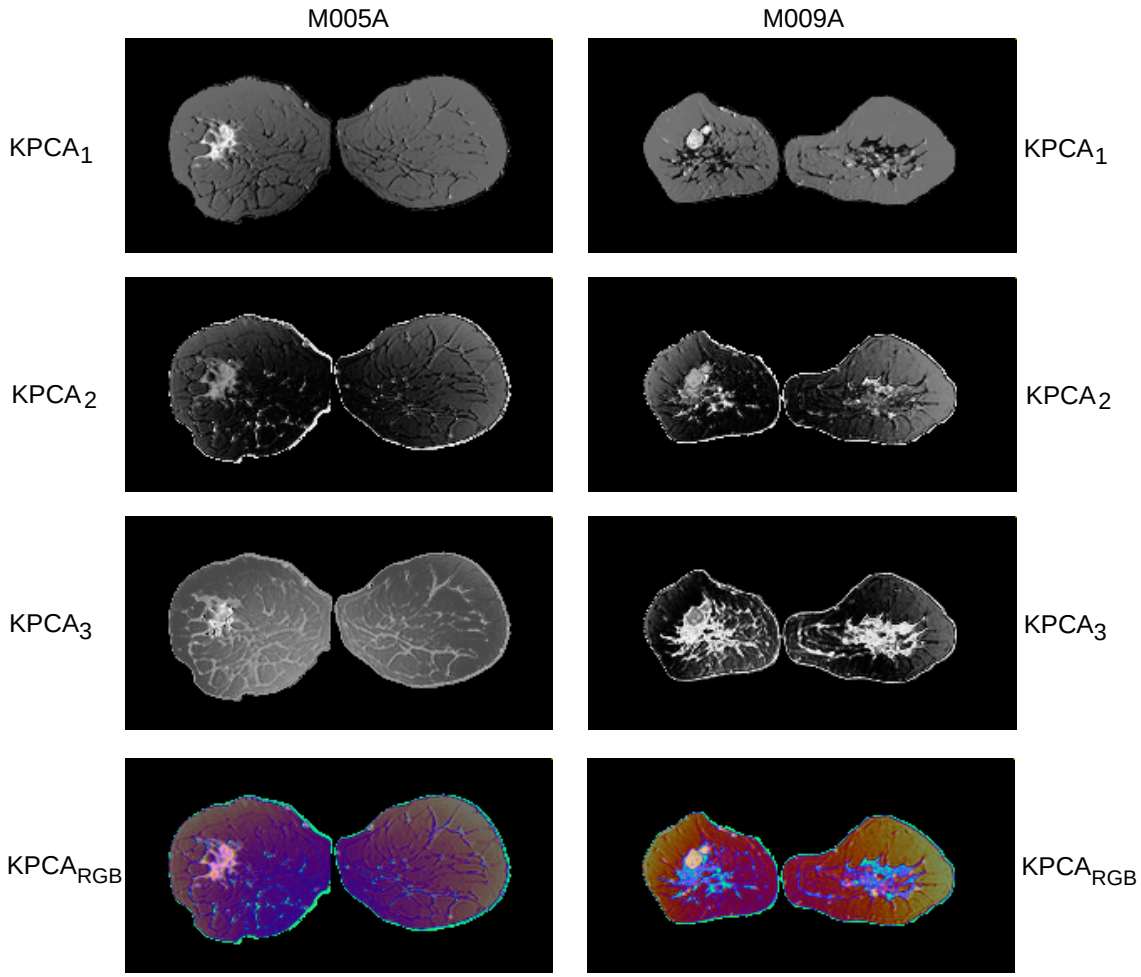


Figure 8.11: Fusion images $\mathrm{KPCA}_1, \ldots, \mathrm{KPCA}_3$ and $\mathrm{KPCA}_{\mathrm{RGB}}$ computed for cases M005A (left column) and M009A (right column). All three principal axes are computed using the same kernel bandwidth. The bandwidth value is chosen according to the maximum of the Fisher's score curve computed for the first principal axis.

with different bandwidth values. The curve of the Fisher's score for $\boldsymbol{\xi}_1$ reaches a maximum for $\sigma \in [0.2; 0.3]$, while the curve of $\boldsymbol{\xi}_2$ has its minimum at this point. According to the Fisher's score, normal and suspicious signals are best separated by $\boldsymbol{\xi}_3$ for a kernel bandwidth $\sigma \in [0.5; 0.6]$.

The four fusion images of the excluded case M005A displayed below the plot illustrate that the Fisher's score is a reasonable criterium for tuning the presentation of lesion masses. The fusion image KPCA_1 computed with the bandwidth value yielding the maximum Fisher's score ($\sigma = 0.2$, Bandwidth B) depicts the lesion mass and the surrounding normal tissue with high visual contrast, i.e. with significantly different intensity levels. This visual contrast decreases for increasing values of σ (Bandwidths C,D), which is also reflected by the decreasing Fisher's score.

The kernel bandwidth σ used for computing the principal axes $\boldsymbol{\xi}_2$ and $\boldsymbol{\xi}_3$ can either be selected dependently or independently from the kernel bandwidth used for computing $\boldsymbol{\xi}_1$. In figure

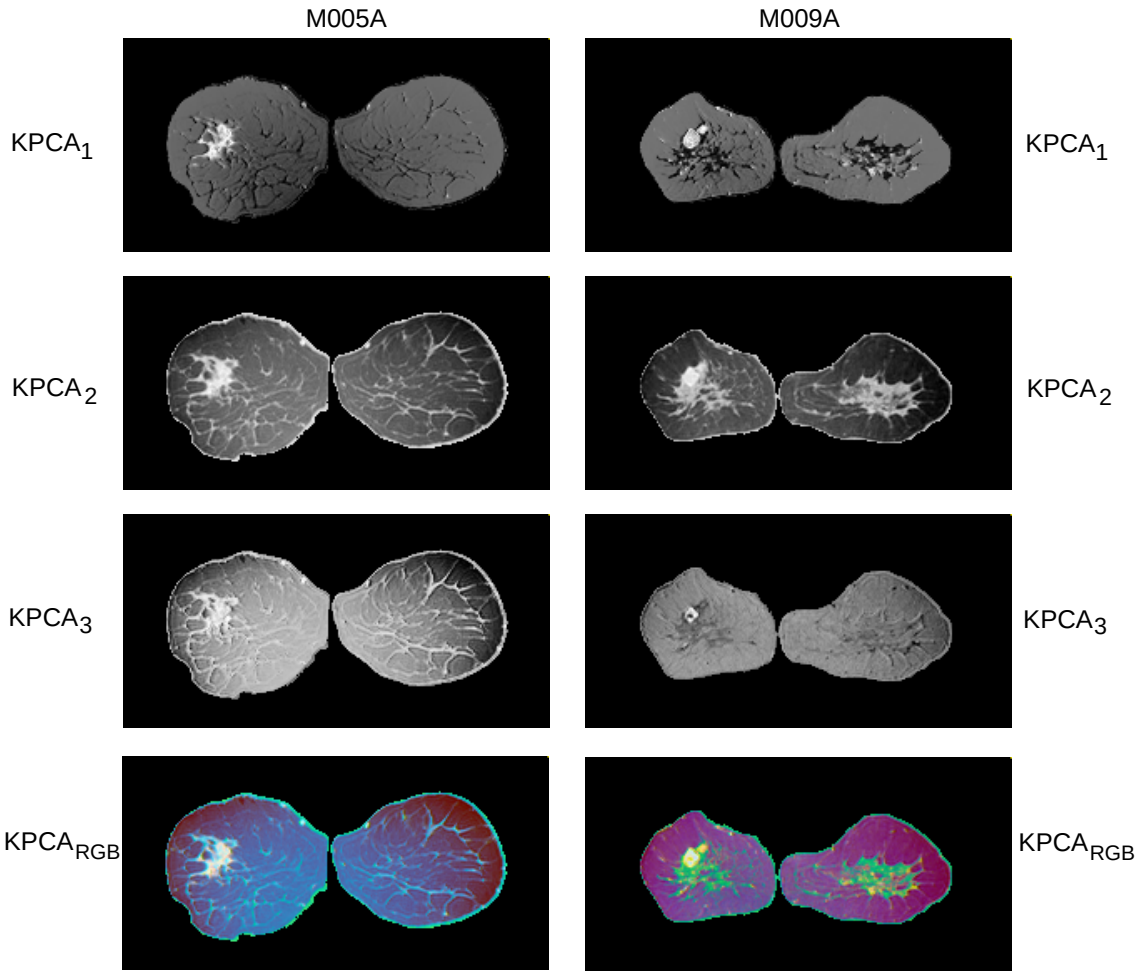


Figure 8.12: Fusion images $\mathrm{KPCA}_1, \ldots, \mathrm{KPCA}_3$ and $\mathrm{KPCA}_{\mathrm{RGB}}$ computed for cases M005A (left column) and M009A (right column). All three principal axes are computed using a different kernel bandwidth. The bandwidth value is chosen according to the maximum of the Fisher's score individually computed for each principal axis on a subset of the training data.

8.11, the principal axes $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_3$, i.e. the fusion images $\mathrm{KPCA}_1, \ldots, \mathrm{KPCA}_3$ and $\mathrm{KPCA}_{\mathrm{RGB}}$, are determined using the same kernel bandwidth for all three axes. Thus, all three principal axes are computed within the same kernel-induced feature space, leading to uncorrelated principal component values. The value of σ was chosen according to the maximum of the course of the Fisher's score computed for $\boldsymbol{\xi}_1$. Hence, the lesions are conspicuously displayed in KPCA_1. Since the maximum of the Fisher's score course computed for $\boldsymbol{\xi}_1$ does not necessarily coincide with the maximum determined for $\boldsymbol{\xi}_2$ and $\boldsymbol{\xi}_3$, the corresponding fusion images KPCA_2 and KPCA_3 may provide a less conspicuous visualisation of the lesions. As illustrated in figure 8.10 for case M005A, the optimal value for the bandwidth of $\boldsymbol{\xi}_1$ may even coincide with the minimum of the Fisher's score for $\boldsymbol{\xi}_2$ and with a low Fisher's score for $\boldsymbol{\xi}_3$, leading to a less prominent depiction of the lesion mass in the fusion images KPCA_2, KPCA_3 and $\mathrm{KPCA}_{\mathrm{RGB}}$ as depicted in figure 8.11.

Table 8.1: Az for fusion images PCA_1, PCA_2 and PCA_3 based on case-specific representation spaces spanned by the first three principal axes of PCA.

Case ID	PCA
M005A	0.547/0.935/0.759
M007A	0.653/0.902/0.684
M009A	0.781/0.904/0.829
B015A	0.702/0.945/0.781
M094A	0.820/0.980/0.625
M104A	0.810/0.860/0.856

In this case, $\boldsymbol{\xi}_1$ lies in the direction which is most sensitive to the characteristic signal courses of suspicious tissue, whereas the orthogonality constraint causes the following principal axes to be less sensitive to lesion signals. Therewith, the lesions are predominantly encoloured with shadings of red in the composite images $KPCA_{RGB}$.

In order to enhance the display of lesion masses in the fusion images $KPCA_2$ and $KPCA_3$, the corresponding bandwidth values can be chosen independently for all three principal axes. In the case that all three principal axes are computed for different values of σ, they relate to different feature spaces $\mathcal{F}$. Therewith, they are not orthogonal to each other and the intensity values of the different fusion images may be correlated and reflect redundant information. Figure 8.12 presents the same coronal slices as figure 8.11 but fused by projecting the corresponding temporal kinetic signals onto individually tuned principal axes. Each principal axis was computed by KPCA with the kernel bandwidth value for which the corresponding Fisher's score curve reaches its maximum. In all three fusion images $KPCA_1, \ldots, KPCA_3$, lesions are displayed as bright masses with medium to high visual contrast to normal tissue. The fusion images $KPCA_2$ and $KPCA_3$ of case M005A exhibit similar image characteristics and represent redundant information. In $KPCA_2$ of M009A, the lesion mass as well as glandular tissue can be identified, whereas $KPCA_1$ depicts the lesion and fat tissue. Due to the image characteristics of $KPCA_1, \ldots, KPCA_3$, the two lesion masses are encoloured intense white in the corresponding composite image $KPCA_{RGB}$. Furthermore, structures such as blood vessels or glandular tissue are depicted more clearly than in the colour images computed by KPCA with a single bandwidth value (Fig. 8.11).

8.3.4 ROC Analysis

In order to analyse to what extent fusion of DCE-MRI sequences can serve as a tool for detecting lesion masses, the fusion images based on the first three principal axes derived from PCA and KPCA are evaluated by measuring the *area-under-the-ROC-curve* (Az). Table 8.1 contains the Az for the fusion images $PCA_1, \ldots, PCA_3$ based on data projections into case-specific representation spaces determined by PCA (setup I).

The Az values for fusion images based on domain-specific representation spaces (setup II) are listed in table 8.2. The first column of table 8.2 lists the results for linear PCA. The second column contains the outcome for KPCA with a single bandwidth value σ selected according to the maximum of the Fisher's score computed for principal components based on the first principal axis. Results for KPCA using an individual bandwidth value for each principal axis can be observed in the fourth column. For comparison, the average Az values yielded by the supervised SVM

Table 8.2: Az for fusion images based on the domain-specific representation space spanned by the first three principal axes computed by PCA (second column), KPCA with a single bandwidth value (third column) and KPCA with three bandwidth values individually optimised for each principal axis. The fourth column reflects the mean Az as computed in chapter 5 (SVM with Gaussian kernel, *raw*-feature, filtered image data).

Case ID	PCA	KPCA (single σ)	KPCA	SVM-G
M005A	0.541/0.938/0.519	0.765/0.616/0.770	0.765/0.809/0.517	0.974
M007A	0.642/0.909/0.499	0.754/0.437/0.728	0.754/0.730/0.826	0.978
M009A	0.753/0.924/0.514	0.827/0.443/0.643	0.827/0.737/0.385	0.977
B015A	0.684/0.937/0.933	0.810/0.605/0.576	0.810/0.926/0.704	0.938
M094A	0.808/0.980/0.559	0.927/0.434/0.675	0.927/0.858/0.671	0.980
M104A	0.749/0.907/0.554	0.719/0.744/0.814	0.719/0.744/0.836	0.935

algorithm with Gaussian kernel evaluating the temporal kinetic signals of mean filtered image sequences (see chapter 5) are listed in the last column.

According to the presented ROC indices, fusion images computed by projecting the image data into a domain-specific representation space based on PCA are most suitable for detecting lesions, although being inferior to the confidence volumes returned by the setup based on supervised classification algorithms[2].

8.4 Discussion

The results presented in this chapter exemplify the value of image fusion for the analysis of DCE-MRI data. By PCA and KPCA of the temporal kinetic signals, the number of images of a DCE-MRI sequence can be reduced to two or three images without losing essential information. Lesions can be discovered by means of examination of the fusion images based on the first and second principal axis. For one case, analysis of the fusion images based on the following principal axes allows for distinguishing lesion compartments with different signal characteristics. The ROC analysis of the considered cases indicates that in particular the fusion image PCA_2 computed by projecting data into a domain-specific representation space (setup II) determined by linear PCA yields high Az values. Similar Az values can be measured for the fusion image PCA_2, if the projection is determined by PCA of the DCE-MRI sequence which is to be fused subsequently (setup I). Although KPCA allows to capture nonlinear correlations between signal components and is therefore more powerful than PCA, ROC analysis indicates an inferior performance for the corresponding fusion images from the viewpoint of lesion detection.

From the viewpoint of data visualisation, the following characteristics can be observed for fusion images based on case-specific representation spaces (setup I): The fusion image PCA_1 based on the principal axis with the largest eigenvalue displays the temporal average of the signal components. The corresponding principal axis exposes a nearly constant or weakly enhancing signal which is typical for normal tissues such as fat, glandular and connectivity tissues. Except of fat, such tissue types are displayed with low intensities in the precontrast image of T1-weighted

[2]Note that the training data of the supervised classifier included two more cases, in particular case B013A whose lesion is difficult to recognise by means of its temporal kinetic signals.

sequences [Heywang-Köbrunner and Beck, 1995]. Thus, the major part of the total variance of DCE-MRI data is caused by signals of non-enhancing tissue types displayed with different intensities in the original image sequence. The second and third principal axes are likely to account for variance caused by signals of tissue affected by pathological disorders. Regions of tissue exposing suspicious temporal kinetic signals are accentuated in the corresponding fusion images. In general, fusion images based on the remaining principal axes provide negligible amounts of information to the observer.

Fusion images based on a domain-specific representation space spanned by the principal axes obtained from PCA exhibit similar characteristics to those computed in setup I. Fusion images based on the first principal axis display the entire breast without accentuating lesions. These are conspicuously depicted in fusion images based on the second and, for some cases, in the fusion images based third principal axis. The corresponding composite images $\mathsf{PCA}_{\mathsf{RGB}}$ allow for localising lesion masses, but the visual contrast between the different types of normal tissues is low compared to the colour images based on KPCA. In particular the fusion images $\mathsf{KPCA}_{\mathsf{RGB}}$ based on individually tuned principal axes display lesions and different types of normal tissue with high visual contrast. Even though the gray value images $\mathsf{KPCA}_1, \ldots, \mathsf{KPCA}_3$ mapped to the red, green and blue component of these composite images yield only medium Az values in the ROC analysis, suggesting that they are less suitable for the purpose of lesion detection, lesions are silhouetted against surrounding tissue and can easily be discerned in the three gray value images as well as in the composite image $\mathsf{KPCA}_{\mathsf{RGB}}$. Furthermore, glandular tissue is depicted in $\mathsf{KPCA}_{\mathsf{RGB}}$ with superior visual contrast to fat tissue than in the composite images based on linear PCA. Thus, composite images based on KPCA with individually tuned principal axes provide the most comprehensive visualisations of DCE-MRI sequences.

A disadvantage of case-specific representation spaces is that the image characteristics of the fusion images may vary for different cases. Since PCA is recomputed for each new case, the information reflected by fusion images based on the first principal axes may change with the size and type of lesions or if cases exhibit no lesion at all. Furthermore, the computation of principal axes using the data of unseen, subsequently fused DCE-MRI sequences is sensitive to motion artefacts. Even spatially restricted motion artefacts may affect a number of temporal kinetic signals which quickly exceed the number of signals caused by pathologically affected tissue. In this case, the principal axes corresponding to the leading eigenvalues are more likely to account for variance caused by the motion artefact than for variance caused by pathological disorders.

An aspect which has to be considered for practical applications is the computational expense of both methods. Solving the eigenvalue problem is of $\mathcal{O}(n_t^3)$ complexity for PCA and of $\mathcal{O}(N^3)$ complexity for KPCA, with n_t being the dimensionality of the data space and N being the number of training examples. In practice, only a few of the N possible principal axes of KPCA are needed. Therefore, the computational expense can greatly be reduced by solving the eigenvalue problem using an iterative algorithms which allows to determine subsets of eigenvectors. The memory complexity is about of $\mathcal{O}(n_t^2)$ for PCA and $\mathcal{O}(N^2)$ for KPCA if the algorithm used for computing the eigenvectors requires the entire covariance or kernel matrix to be stored in memory. The computational complexity of a single principal component scales linearly with the number of training examples for KPCA and linearly with the data dimension for PCA. Thus, the size N of the training set Γ_{Train} significantly influences the computational expenses for computing principal axes but also, and more important for practical applications, for fusing DCE-MRI sequences by

KPCA. Computing the fusion images based on the first three principal axes takes approximately one minute per case for PCA but 25min per case for KPCA (Pentium IV, 1.8GHz, 2GB RAM, $N = 1000$).

In summary, both setups allow for representing DCE-MRI sequences by a reduced number of images which, nevertheless, visually emphasise tissue structures that are relevant for breast cancer diagnosis. According to the ROC analysis, mapping of temporal kinetic signals into a domain-specific representation space spanned by PCA leads to fusion images which are particular suitable for detecting lesions. The RGB composite images obtained from image fusion based on a domain-specific representation space determined by KPCA allow for examination of the structure of the entire breast and enable radiologists to distinguish healthy and pathologically affected tissue as well as different types of normal tissue by means of a single colour image. Therewith, fusion images computed by PCA or KPCA of the temporal kinetic signals suffice the two objectives *lesion detection* and *compact visualisation of entire image volumes*. To some extent, the fusion images also expose further information about the internal structure of lesions and therewith suffice the third objective *characterisation of lesion tissue*.

9 Conclusion

The aim of this thesis was to investigate the utility of data-driven analysis of *dynamic contrast-enhanced magnetic resonance imaging* (DCE-MRI) data recorded for the purpose of breast cancer diagnosis. In classical image processing, the spatial order of signal values receives much attention. Since a single pixel or voxel of an univariate image provides only little information, the local informational content is frequently evaluated by means of features reflecting the *spatial signal characteristics* of small image windows as quantified by e.g. first and second order statistics of gray values. Thus, a true voxel-by-voxel evaluation of such images is often not possible, since the corresponding feature vectors do not solely relate to single voxels but to small image patches depicting their neighbourhoods. In the end, this causes a certain spatial imprecision of the outcome obtained from evaluations of such feature vectors. In contrast, voxels of multivariate images like DCE-MRI images possess much more information, and the temporal kinetic signals on their own as measured by DCE-MRI have demonstrated to provide sufficient information for discerning different types of tissue. Therewith, true voxel-by-voxel processing of such image data becomes possible.

The new methods as proposed within this work are predominantly focused on evaluating the temporal component of DCE-MRI data as given by temporal kinetic signals of individual voxels, enabling physicians to analyse tissue disorders with the same spatial resolution as the image data themselves. To this end, image sequences are regarded as data distributions in a signal space, which are subject of analysis by supervised and unsupervised artificial neural networks and machine learning algorithms. The proposed analysis modules have in common that they are **model-free** and **data-driven**; physicians are neither required to predefine concrete attributes in which signals of different tissue classes vary nor to design an explicit mathematical model of the physiological process underlying the signal. Instead, ANNs were applied to derive such distinguishing signal characteristics autonomously from the data themselves by correlating statistical features of the image data with supplemental information. Such supplemental information can be the outcome of a manual image evaluation, e.g. lesion labels as used in chapter 5 for whose generation a radiologist adopted his primary medical expertise, or information obtained from other routinely applied diagnosis techniques such as the outcome of histological examinations of tissue samples as described in chapter 6. Even without any side information, hidden regularities in the signal data were revealed by applying unsupervised ANNs. Therewith, data-driven approaches as presented throughout this work are advantageous whenever a-priori knowledge about signal characteristics is limited or the physiological process underlying the signal is too complex to be adequately modelled.

Each module introduced in this work supports radiologists at certain steps of DCE-MRI analysis. Thereby, **each module suffices one or several of the three main areas of applications of computer aided diagnosis (CAD) systems in DCE-MRI** as outlined in chapter 3.4: Confidence volumes computed with state-of-the-art supervised machine learning algorithms, correlating expert labels of lesion masses with DCE-MRI data as described in chapter 5, allow to **localise**

lesions and to delineate their extent by means of visual examination of a single three-dimensional image per case. The subsequent **detailed characterisation of lesion masses** is supported by augmenting conventional DCE-MRI data visualisations with pseudo-colours as described in chapter 6. For this purpose, supervised ANNs were employed to correlate signal features with class information derived from manual lesion segmentations and histological examinations of tissue samples. The pseudo-colours obtained from this pixel-mapping function reflect the probabilities that the evaluated signals were caused by normal, malignant or benign tissue and allow to visualise temporal characteristics of signals and their spatial distribution by means of single colour images. A pseudo-colour image of the entire field-of-view or a selective colouring of subregions suspected to display tissue disorders exposes the distributions of the different signal classes in the heterogenous lesion tissue. Thus, lesion compartments with homogenous signal characteristics can be designated in order to measure average signals of ROIs more reliably. Furthermore, the visualisation of the distribution of benign and malignant signals itself provides valuable information for the evaluation of morphological criteria which are, beside the temporal characteristics, essential for classifying lesions. Image fusion methods as presented in chapter 8 allow to compute **compact visualisations of entire image volumes**. The individual images of a DCE-MRI sequence were combined to new images by solely evaluating statistical properties of the DCE-MRI data. Thereby, the number of images that are to be analysed by radiologists was reduced, and phenomenons such as tissue disorders, which are inherent in the image data but not evident in conventional images, were accentuated.

The outcome of each module was carefully evaluated by ROC analysis, which is the current standard in biomedical engineering, but also by a detailed discussion of its visual presentation. The qualitative evaluation of the visualisations has shown to be of particular importance for two reasons. Firstly, ROC analysis regards the outcome from a pure statistical viewpoint by only evaluating the rank order of confidence values assigned to individual voxels. But the influence on the diagnosis of spuriously highlighted voxels or *false positives* also strongly depends on their spatial arrangements. Isolated false positives scattered over the entire image volume or clusters of such voxels located in regions of the heart cavity, which are frequently caused by the high concentration of contrast agent in the blood, can be regarded to be less misleading to the observer than clusters of false positives located in the breast. The latter have to be regarded as potential lesions whereas the former can be identified as falsely highlighted by the virtue of their spatial locations in the thorax. Secondly, in particular in biomedical domains, it is often difficult or even impossible to acquire definite reference labels which perfectly reflect the biological truth. For instance, the lesion label as utilised in this work does not primarily reflect the classification of individual signals but the radiologist's subjective assessment of the entire lesion for which he also considers morphological criteria or the lesion's appearance in images obtained from complementary imaging modalities. Thus, this label does not exactly provide the information that is needed for a ROC analysis of confidence values, reflecting the assessment of individual temporal kinetic signals. Furthermore, the lesion assessment strongly depends on the radiologist's expertise and is often affected by inter- and intra-observer variance. In this situation, the qualitative and quantitative comparison with established clinical DCE-MRI analysis techniques such as subtraction images or the 3TP method has shown to be helpful for illustrating and analysing how the data-driven approaches perform in comparison to other standard techniques solely evaluating signals of individual voxels.

9.1 Subsequent Steps in Data-Driven Analysis of DCE-MRI Data

The modules presented in this thesis support radiologists throughout the analysis of DCE-MRI data. The modules purify the information needed for the different working steps and give a visual feedback that enables physicians to localise lesion masses or to visually analyse their architecture. However, the final classification of lesion masses still requires the radiologist's expertise.

The next step in data-driven diagnosis of breast cancer by means of DCE-MRI is the automatic classification of entire lesion masses. For this purpose, it will be essential to regard lesions as a whole, i.e. to correlate the temporal and spatial appearance of the entire conglomerate of lesion voxels. In particular, quantification of the morphology of lesions remains an open question but has been stated by many physicians to be a key component for a reliable classification of breast cancer lesions. A first approach would be to regard lesion masses as unstructured sets of temporal kinetic signals. Each lesion would form an individual distribution in the signal space, and the similarity of entire lesions could be compared using suitable metrics such as the *Earth Mover's distance* [Rubner et al., 2000]. Since morphology, i.e the spatial arrangements of signals, plays a crucial role in classification of lesions, this approach should then be extended in such a way that the spatial order of the temporal kinetic signals is considered. Presently, the processing of such structured data also gains much attention from the machine learning community. Kernel-based methods offer several appealing properties which make them particularly suitable for processing structured input data, and several kernel functions for sets of vectors or data structured as e.g. trees have been proposed. Yet, a prerequisite for data-driven analysis of entire lesion will be the availability of image data for a large number of cases.

Another interesting extension arises from the fact that during clinical diagnosis processes, tissue, in particular lesion tissue, is commonly investigated from a variety of viewpoints. Diagnosis techniques such as DCE-MRI, X-ray, histopathological examination or gene-profiling provide a broad range of complementary information exposing macroscopic, microscopic or even subcellular tissue features. An important question is, therefore, how such heterogenous types of information can be processed in a unifying approach. Unlike model-based approaches, artificial neural networks are able to evaluate feature vectors containing heterogenous types of features. Thus, the data-driven approaches as presented in this thesis can easily be extended to processing input patterns consisting of features derived from DCE-MRI sequences and e.g. features reflecting tissue parameters recorded by hyperspectral MRI, given that both types of image data have been coregistered. The challenge will be to align image data of different dimensionalities and resolutions and, most notably, to align image data with nonimage data. Artificial neural networks and machine learning algorithms have been shown to be promising tools for processing of multivariate data. In particular in medical research, explorative approaches such as data-driven techniques are particularly suitable for data analysis. During exploration of unknown diseases and development of new techniques for treatment, complex data will be acquired from heterogeneous sources but exact knowledge about how certain disorders are reflected by the different types of data will often be limited at that stage. In this situation, data-driven techniques will be adjuvant tools to validate hypotheses derived from empirical observations or even to reveal unknown relationships between different data components which are inherent in the data but not apparent to the physicians.

Furthermore, they will allow to incorporate existing knowledge into the analysis process in order to improve their efficiency. Thus, data-driven analysis of medical data based on artificial neural networks and machine learning should always be accompanied by a dialog between computer scientists and physicians in order to find an appropriate preprocessing of data, to design reasonable setups of algorithms, to discuss the outcome regarding its plausibility and, finally, to derive new medical knowledge.

Bibliography

[Abdolmaleki et al., 2001] Abdolmaleki, P., Buadu, L. D., and Naderimansh, H. (2001). Feature extraction and classification of breast cancer on dynamic magnetic resonance imaging using artificial neural network. *Cancer Letters*, 171:183–191.

[Aizerman et al., 1964] Aizerman, M., Braverman, E., and Rozonoer, L. (1964). Theoretical foundations of the potential function method in pattern recognition learning. *Automation and remote control*, 25:821–837.

[Aldroubi and Unser, 1996] Aldroubi, A. and Unser, M. (1996). *Wavelets in medicine and biology*. CRC Press.

[Allwein et al., 2000] Allwein, E. L., Schapire, R. E., and Singer, Y. (2000). Reducing multiclass to binary: A unifying approach for margin classifiers. *Journal of Machine Learning Research*, 1:113–141.

[Bankman, 2000] Bankman, I. N., editor (2000). *Handbook of medical imaging*. Academic Press.

[Bennett and Campbell, 2000] Bennett, K. P. and Campbell, C. (2000). Support vector machines: hype or hallelujah? *SIGKDD Exploration Newsletter*, 2(2):1–13.

[Bishop, 1995] Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford University Press.

[Bradley, 1997] Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7):1145–1159.

[Breimann et al., 1998] Breimann, L., Friedmann, J., Olshen, R., and Stone, C. (1998). *Classification and regression trees*. Chapman & Hall.

[Brown et al., 2000] Brown, J., Buckley, D., Coulthard, A., Dixon, A. K., Dixon, J. M., Easton, D. F., Eeles, R. A., Evans, D. G., Gilbert, F. G., Graves, H., Hayes, C., Jenkins, J., Jones, A. P., Keevil, S. F., Leach, M. O., Liney, G. P., Moss, S. M., Padhani, A. R., Parker, G. J., Pointon, L. J., Ponder, B. A., Redpath, T. W., Sloane, J. P., Turnbull, L. W., Walker, L. G., and Warren, R. M. (2000). Magnetic resonance imaging screening in women at genetic risk of breast cancer: imaging and analysis protocol for the uk multicentre study. UK MRI breast screening study advisory group. *Magnetic Resonance Imaging*, 18(7):765–76.

[Castellano et al., 2004] Castellano, G., Bonilha, L., Li, L. M., and Cendes, F. (2004). Texture analysis of medical images. *Clinical Radiology*, 59(1061–1069).

[Chapelle et al., 2002] Chapelle, O., Vapnik, V., Bousquet, O., and Mukherjee, S. (2002). Choosing multiple parameters for support vector machines. *Machine Learning*, 46(1-3):131–159.

[Chen et al., 2004] Chen, W., Giger, M. L., and Bick, U. (2004). Automated identification of temporal pattern with high initial enhancement in dynamic MR lesions using fuzzy c-means algorithm. In Fitzpatrick, J. M. and Sonka, M., editors, *Proceedings of SPIE 2004*, volume 5370.

[Collins and Padhani, 2004] Collins, D. J. and Padhani, A. R. (2004). Dynamic magnetic resonance imaging of tumor perfusion. *IEEE Engineering in Medicine and Biology Magazine*, 23(5):65–83.

[Cristianini and Shawe-Taylor, 2000] Cristianini, N. C. and Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press.

[Daubechies, 1991] Daubechies, I. (1991). *Ten lectures on wavelets*. CBMS-SFS Series Applied Mathematics, SIAM.

[Dean and Rodrick, 1983] Dean, S. R. and Rodrick, S. (1983). *The Radon transform and some of its applications*. Wiley.

[DeLong et al., 1988] DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristics curves: A nonparametric approach. *Biometrics*, 44:837–845.

[DICOM14, 2004] DICOM14, N. E. M. A. (2004). Digital imaging and communications in medicine (DICOM), part 14: Grayscale standard display function.

[Dietterich and Bakiri, 1995] Dietterich, T. G. and Bakiri, G. (1995). Solving multiclass learning problems via error-correcting output codes. *Journal of Artificial Intelligence Research*, 2:263–286.

[Duan et al., 2003a] Duan, K., Keerthi, S. S., and Poo, A. N. (2003a). Evaluation of simple performance measures for tuning SVM hyperparameters. *Neurocomputing*, 51:41–59.

[Duan et al., 2003b] Duan, K., Keerthi, S. S., Shevade, S. K., Chu, W., and Poo, A. N. (2003b). Multi-category classification by soft-max combination of binary classifiers. In *4th International Workshop on Multiple Classifier Systems*.

[Engelmeier et al., 2004] Engelmeier, K. H., Hellwig, G., Griebel, J., Delorme, S., Siebert, M., and Brix, G. (2004). Morpho-functional visualization of dynamic MR-mammography. In Fieschi, M., editor, *Proceedings of MEDINFO 2004*.

[Eustat, 2004] Eustat (2004). *Regions: Statistical yearbook 2004*. European Commission.

[Fawcett, 2003] Fawcett, T. (2003). ROC graphs: Notes and practical considerations for researchers. Technical report, HP Labs.

[Fiori and Piazza, 1999] Fiori, S. and Piazza, F. (1999). A comparision of three PCA neural techniques. In *Proceedings of the European Symposium of Neural Networks*.

[Flach, 2003] Flach, P. A. (2003). The geometry of ROC space: Understanding machine learning metrics through ROC isometrics. In *Proceeddings of the 20th International Conference on Machine Learning*, pages 194–201. AAAI Press.

[Fleiss, 1982] Fleiss, J. L. (1982). *Statistical methods for rates and proportions*. Wiley.

[Fogel and Sagi, 1989] Fogel, I. and Sagi, D. (1989). Gabor filters as texture discriminator. *Biological Cybernetics*, 10:103–113.

[Friedman and Tukey, 1974] Friedman, J. H. and Tukey, J. W. (1974). A projection pursuit algorithm for exploratory data analysis. *IEEE Transactions on Computers*, 23(9):881–890.

[Friedrichs and Igel, 2004] Friedrichs, F. and Igel, C. (2004). Evolutionary tuning of multiple SVM parameters. In Verleysen, M., editor, *12th European Symposium on Artificial Neural Networks (ESANN 2004)*, pages 519–524.

[Fukunaga, 1990] Fukunaga, K. (1990). *Statistical pattern 2ecognition*. Academic Press.

[Gibbs and Turnball, 2003] Gibbs, P. and Turnball, L. W. (2003). Textural analysis of contrast-enhanced MR images of the breast. *Magnetic Resonance in Medicine*, 50:92–98.

[Grigorescu et al., 2002] Grigorescu, S. E., Petkov, N., and Kruizinga, P. (2002). Comparison of texture features based on Gabor filters. *IEEE Transactions on Image Processing*, 11(10):1160–1167.

[Hamm et al., 1994] Hamm, B., Thoeni, R. F., Gould, R. G., Bernardino, M. E., Luning, M., Saini, S., Mahfouz, A. E., Taupitz, M., and Wolf, K. J. (1994). Focal liver lesions: Characterization with nonenhanced and dynamic contrast material-enhanced MR imaging. *Radiology*, 190:417–423.

[Hanley, 1989] Hanley, J. A. (1989). Receiver operating characteristic methodology: The state of the art. *CRC critical reviews in diagnostic imaging*, 29:307–335.

[Hanley and McNeil, 1982] Hanley, J. A. and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristics (ROC) curve. *Radiology*, 143:29–36.

[Hanley and McNeil, 1983] Hanley, J. A. and McNeil, B. J. (1983). A method of comparing the areas under receiver operating charcterisitc curves derived from the same cases. *Radiology*, 148:839–843.

[Haralick et al., 1973] Haralick, R. M., Shanmugam, K., and Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, 3(6):610–621.

[Hastie and Tibshirani, 1998] Hastie, T. and Tibshirani, R. (1998). Classification by pairwise coupling. In Jordan, M. I., Kearns, M. J., and Solla, S. A., editors, *Advances in Neural Information Processing Systems*, volume 10. The MIT Press.

[Hastie et al., 2001] Hastie, T., Tibshirani, R., and Friedmann, J. (2001). *The elements of statistical learning*. Springer.

[Hellwig et al., 2002] Hellwig, G., Brix, G., Griebel, J., Lucht, R., Delorme, S., Siebert, M., and Engelmeier, K. H. (2002). Dynamic MR mammography: Three-dimensional real-time visualization of contrast enhancement in virtual reality. *Academic Radiololgy*, 9:1255–1263.

[Herlidou-Meme et al., 2003] Herlidou-Meme, S., Constans, J. M., Carsin, B., Olivie, D., Eliat, P. A., Nadal-Desbarats, L., Gondry, C., Rumeur, E. L., Idy-Peretti, I., and de Certaines, J. D. (2003). MRI texture analysis on texture test objects, normal brain and intracranial tumors. *Magnetic Resonance Imaging*, 21(9):989–993.

[Heywang-Köbrunner and Beck, 1995] Heywang-Köbrunner, S. H. and Beck, R. (1995). *Contrast-enhanced MRI of the breast*. Springer.

[Holden et al., 1995] Holden, M., Steen, E., and Lundervold, A. (1995). Segmentation and visualization of brain lesions in multispectral magnetic resonance images. *Computerised Medical Imaging and Graphics*, 19(2):171–83.

[Hornak, 2004] Hornak, J. P. (2004). The basics of MRI. Online http://www. cis. rit. edu/htbooks/mri/index. html.

[Hsu and Lin, 2002] Hsu, C. and Lin, C. (2002). A comparison of methods for multi-class support vector machines. *IEEE Transactions on Neural Networks*, 13:415–425.

[Huber, 1985] Huber, P. J. (1985). Projection pursuit. *Annals of Statistics*, 13(2):435–475.

[Hyvärinen et al., 2001] Hyvärinen, A., Karhunen, J., and Oja, E. (2001). *Independent component analysis*. Wiley.

[Hyvärinen and Oja, 2000] Hyvärinen, A. and Oja, E. (2000). Independant component analysis: Algorithms and applications. *Neural Networks*, 13(411-430).

[Jacobs et al., 2003] Jacobs, M. A., Barker, P. B., Bluemke, D. A., Maranto, C., Arnold, C., Herskovites, E. H., and Bhujwalla, Z. (2003). Benign and malignant breast lesions: Diagnosis with multiparametric MR imaging. *Radiology*, 229:225–232.

[Jacobs et al., 1991] Jacobs, R. A., Jordan, M. I., Nowlan, S. J., and Hinton, G. E. (1991). Adaptive mixtures of local experts. *Neural Computation*, 3:79–97.

[James et al., 2001] James, D., Clymer, B. D., and Schmalbrock, P. (2001). Texture detection of simulated microcalcifications susceptibility effects in magnetic resonance imaging of breast. *Journal of Magnetic Resonance Imaging*, 13:876–881.

[Jiang et al., 1996] Jiang, Y., Metz, C. E., and Nishikawa, R. M. (1996). A receiver operating characteristic partial area index for highly sensitive diagnostic tests. *Radiology*, 201(3):745–750.

[Jolliffe, 1986] Jolliffe, I. T. (1986). *Principal component analysis*. Springer.

[Keerthi, 2002] Keerthi, S. S. (2002). Efficient tuning of SVM hyperparameters using radius/margin bound and iterative algorithms. *IEEE Transactions on Neural Networks*, 13:1225–1229.

[Kelcz et al., 2001] Kelcz, F., Furman-Haran, E., Grobgeld, D., and Degani, H. (2001). Clinical testing of high-spatial-resolution parametric contrast-enhanced MR imaging of the breast. *American Journal of Radiology*, 179.

[Kmer et al., 1995] Kmer, K., Ring, P., Thomson, C., and Henriksen, O. (1995). Texture analysis in quantitative MR imaging. *Acta Radiologica*, 36.

[Kohonen, 1995] Kohonen, T. (1995). *Self-organising maps*. Springer.

[Kohonen, 1997] Kohonen, T. (1997). *Self-organizing maps*. Springer.

[Kuhl et al., 1999] Kuhl, C. K., Mielcareck, P., Klaschik, S., Leutner, C., Wardelmann, E., Gieseke, J., and Schild, H. H. (1999). Dynamic breast MR imaging: Are signal intensity time course data useful for differential diagnosis of enhancing lesions? *Radiology*, 211(101).

[Lauterbur, 1973] Lauterbur, P. C. (1973). Image formation by induced local interactions: Examples employing nuclear magnetic resonance. *Nature*, 242:190–191.

[Laws, 1979] Laws, K. I. (1979). Texture energy measures. In *DARPA Image Understanding Workshop 1979*.

[Leach, 2001] Leach, M. O. (2001). Application of magnetic resonance imaging to angiogenesis in breast cancer. *Breast Cancer Research*, 3:22–27.

[Leong et al., 1999] Leong, F. J. W.-M., Graham, A. K., Gahm, T., and McGee, J. O. (1999). Telepathology: Clinical utility and methodology. *Recent Advances in Histopathology*, 18.

[Leslie et al., 2002] Leslie, C., Eskin, E., and Noble, W. S. (2002). The spectrum kernel: A string kernel for SVM protein prediction. In *Proceedings of the Pacific Symposium on Biocomputing 2002*.

[Levkowitz and Herman, 1992] Levkowitz, H. and Herman, G. T. (1992). Color scales for image data. *IEEE Computer Graphics & Applications*, 12(1):72–80.

[Lin et al., 2003] Lin, H., Lin, C., and Weng, R. (2003). A note on Platt's probabilistic outputs for support vector machines.

[Lucht et al., 2002] Lucht, R., Delorme, S., and Brix, G. (2002). Neural network-based segmentation of dynamic MR mammographic images. *Magnetic Resonance Imaging*, 20:147–154.

[Lucht et al., 2001] Lucht, R., Knopp, M. V., and Brix, G. (2001). Classification of signal-time curves from dynamic MR mammography by neural networks. *Magnetic Resonance Imaging*, 19:51–57.

[Manduca, 1994] Manduca, A. (1994). Multi-spectral medical image visualization with self-organizing maps. In *Proceedings ICIP-94 (Cat. No. 94CH35708)*, volume 1, pages 633–7, Los Alamitos, CA, USA. IEEE Comput. Soc. Press.

[Martinetz et al., 1993] Martinetz, T. M., Berkovich, S. G., and Schulten, K. J. (1993). 'Neural-Gas' network for vector quantization and its application to time-series prediction. *IEEE Transactions on Neural Networks*, 4:558–569.

[McClish, 1989] McClish, D. K. (1989). Analyzing a portion of the ROC curve. *Medical Decision Making*, 9(3):190–195.

[Meinicke et al., 2004] Meinicke, P., Tech, M., Morgenstern, B., and Merkl, R. (2004). Oligo kernel for datamining on biological sequences: A case study on prokaryotic translation initiation sites. *BMC Bioinformatics*, 5.

[Metz, 1986] Metz, C. E. (1986). ROC methodology in radiolgic imaging. *Investigative Radiology*, 21:720–733.

[Meyer-Bäse et al., 2004] Meyer-Bäse, A., Wismüller, A., Lange, O., and Leinsinger, G. (2004). Computer-aided diagnosis in breast MRI based on unsupervised clustering techniques. In *Intelligent Computing: Theory and Applications II. Proceedings of the SPIE*, volume 5421, pages 29–37.

[Nason, 1995] Nason, G. (1995). Three-dimensional projection pursuit. *Applied Statistics*, 44(4):411–430.

[Novikoff, 1962] Novikoff, A. B. J. (1962). On convergence proofs on perceptrons. In *Proceedings of the Symposium on the Mathematical Theory of Automata*, volume 12, pages 615–622.

[Obuchowski, 2003] Obuchowski, N. A. (2003). Receiver operating characteristic curves and their use in radiology. *Radiology*, 229:3–8.

[Oja, 1992] Oja, E. (1992). Principal components, minor components, and linear neural networks. *Neural Networks*, 5:927–935.

[Orel, 1999] Orel, S. G. (1999). Differentiating benign from malignant enhancing lesions identified at MR imaging of the breast: Are time-signal intensity curves an accurate predictor? *Radiology*, 211.

[Orel and Schnall, 2001] Orel, S. G. and Schnall, M. D. (2001). MR imaging of the breast for the detection, diagnosis, and staging of breast cancer. *Radiology*, 220(13-30).

[Otsu, 1979] Otsu, N. (1979). A threshold selection method from gray level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1):62–66.

[Piccoli, 1997] Piccoli, C. W. (1997). Contrast-enhanced breast MRI: Factors affecting sensitivity and specificity. *European Radiology*, 7.

[Pizer, 1981] Pizer, S. M. (1981). Intensity mappings to linearize display devices. *Computer Graphics and Image Procesing*, 17(3):262–268.

[Platt, 2000] Platt, J. (2000). Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. In Smola, A. J., Bartlett, P., Schoelkopf, B., and Schuurmans, D., editors, *Advances in Large Margin Classiers*, pages 61–74. MIT Press.

[Platt et al., 2000] Platt, J., Cristianini, N., and Shawe-Taylor, J. (2000). Large margin DAGS for multiclass classification. In Solla, S. A., Leen, T. K., and Mueller, K. R., editors, *Advances in Neural Information Processing Systems 12*, pages 547–553.

[Pohl and van Genderen, 1994] Pohl, C. and van Genderen, J. L. (1994). Image fusion: Issues, techniques and applications. In van Gendern, J. L. and Cappellini, V., editors, *Proceedings EARSeL Workshop*.

[Press et al., 1986] Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1986). *Numerical recipes: The art of scientific computing*. Cambridge University Press, Cambridge (UK) and New York, 1st edition.

[Provost and Fawcett, 2001] Provost, F. and Fawcett, T. (2001). Robust classification for imprecise environments. *Machine Learning Journal*, 42(3).

[Provost et al., 1998] Provost, F., Fawcett, T., and Kohavi, R. (1998). The case against accuracy estimation for comparing induction algorithms. In *Proceedings of the 15th International Conference on Machine Learning*.

[Provost and Fawcett, 1997] Provost, F. J. and Fawcett, T. (1997). Analysis and visualization of classifier performance: Comparison under imprecise class and cost distributions. In *Knowledge Discovery and Data Mining*, pages 43–48.

[Rankin, 2000] Rankin, S. C. (2000). MRI of the breast. *The British Journal of Radiology*, 73:806–818.

[Richards, 1993] Richards, J. A. (1993). *Remote sensing digital image analysis - An introduction*. Springer.

[Richards and Milne, 1983] Richards, J. A. and Milne, A. K. (1983). Mapping fire burns and vegetation regeneration using principal component analysis. In *Proceedings International Geoscience and Remote Sensing Symposium*.

[Ripley, 1996] Ripley, B. D. (1996). *Pattern recognition and neural networks*. Cambridge University Press.

[Ritter, 1991a] Ritter, H. (1991a). Asymptotic level density for a class of vector quantisation processes. *IEEE Transactions on Neural Networks*, 2(1):173–175.

[Ritter, 1991b] Ritter, H. (1991b). Learning with the self-organizing map. In Kohonen, T., editor, *Artificial Neural Networks*, pages 379–384, Amsterdam. Elsevier Science Publishers.

[Ritter et al., 1992] Ritter, H., Martinetz, T., and Schulten, K. (1992). *Neural Computation and Self-Organizing Maps*. Addison-Wesley.

[Rosenblatt, 1958] Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization and brain. *Psychological Review*, 65(6):386–408.

[Rubner et al., 2000] Rubner, Y., Tomasi, C., and Guibas, L. J. (2000). The Earth Mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99–121.

[Saalbach et al., 2004] Saalbach, A., Twellmann, T., Nattkemper, T. W., White, M. J., Khazen, M., and Leach, M. O. (2004). Visualization of multivariate image data using image fusion and perceptually optimized color scales based on sRGB. In Robert L. Galloway, J., editor, *Proceedings of Medical Imaging 2004*, volume 5367, San Diego. SPIE. accepted.

[Sanger, 1989] Sanger, T. D. (1989). Optimal unsupervised learning in a single-layer linear feedforward neural network. *Neural Networks*, 2:459–473.

[Schölkopf et al., 1999a] Schölkopf, B., and C. Burges, S. M., Knirsch, Müller, K. R., Rätsch, G., and Smola, A. J. (1999a). Input space vs. feature space in kernel-based methods. *IEEE Transanctions on Neural Networks*, 10(5):1000–1017.

[Schölkopf and Smola, 2002] Schölkopf, B. and Smola, A. J. (2002). *Learning with kernels*. MIT Press.

[Schölkopf et al., 1998] Schölkopf, B., Smola, A. J., and Müller, K. R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5):1299–1319.

[Schölkopf et al., 1999b] Schölkopf, B., Smola, A. J., and Müller, K. R. (1999b). *Advances in kernel methods - support vector learning*. MIT Press.

[Shawe-Taylor and Cristianini, 2004] Shawe-Taylor, J. and Cristianini, N. (2004). *Kernel methods for pattern analysis*. Cambridge Unversity Press.

[Subramaninan et al., 2004] Subramaninan, K. R., Brockway, J. P., and Carruthers, W. B. (2004). Interactive detection and visualisation of breast lesions from dynamic enhanced MRI volumes. *Computerized Medical Imaging and Graphics*. in press.

[Swets and Pickett, 1982] Swets, J. A. and Pickett, R. M. (1982). *Evaluation of dignostic systems*. Academic Press.

[Szabo et al., 2003] Szabo, B. K., Aspelin, P., Wiberg, M. K., and Bone, B. (2003). Dynamic MR imaging of the breast - analysis of kinetic and morphologic diagnostic criteria. *Acta Radiologica*, 44:379–386.

[Thikhonov and Arsenin, 1977] Thikhonov, A. N. and Arsenin, V. Y. (1977). *Solutions of ill-posed problems*. Wiley.

[Tofts, 1997] Tofts, P. S. (1997). Modelling tracer kinetics in dynamic Gd-DTPA MR imaging. *Journal of Magnetic Resonance Imaging*, 7:91–101.

[Tofts et al., 1999] Tofts, P. S., Brix, G., Buckley, D. L., Evelhoch, J. L., Henderson, E., Knopp, M. V., Larrson, H. B. W., Lee, T. Y., Mayr, N. A., Parker, G. J. M., Port, R. E., Taylor, J., and Weisskopf, R. M. (1999). Estimating kinetic parameters from dynamic contrast enhanced t1-weighted MRI of a diffusible tracer: Standarized quantities and symbols. *Journal of Magnetic Resonance Imaging*, 10:223–232.

[Tofts and Kermode, 1991] Tofts, P. S. and Kermode, A. G. (1991). Measurement of the blood-brain barrier permeability and leakage space using dynamic MR imaging. *Magnetic Resonance In Medicine*, 17:357–367.

[Tyczynski et al., 2002] Tyczynski, J. E., Bray, F., and Parkin, D. M. (2002). Breast cancer in Europe. *ENCR Cancer Fact Sheets*, 2.

[Tzacheva et al., 2003] Tzacheva, A. A., Najarian, K., and Brockway, J. P. (2003). Breast cancer detection in Gadolinium enhanced MR images by static region descriptors and neural networks. *Journal of Magnetic Resonance Imaging*, 17(3):337–342.

[Vaidyanathan et al., 1995] Vaidyanathan, M., Clarke, L. P., Velthuizen, R. P., Phuphanich, S., Bensaid, A. M., Hall, L. O., Bezdek, J. C., Greenberg, H., Trotti, A., and Silbiger, M. (1995). Comparison of supervised MRI segmentation methods for tumor volume determination during therapy. *Magnetic Resonance Imaging*, 13(5):719–28.

[Valdes-Cristerna et al., 2004] Valdes-Cristerna, R., Medina-Banuelos, V., and Yanez-Suarez, O. (2004). Coupling of radial-basis network and active contour models for multispectral brain MRI segmentations. *IEEE Transactions on Biomedical Enginerring*, 51(3):459–470.

[Vapnik, 1995] Vapnik, V. (1995). *The nature of statistical learning theory*. Springer.

[Vesanto and Alhoniemi, 2000] Vesanto, J. and Alhoniemi, E. (2000). Clustering of self-organising map. *IEEE Transactions on Neural Networks*, 11(3):586–600.

[Villmann et al., 2003] Villmann, T., Merenyi, E., and Hammer, B. (2003). Neural maps in remote sensing image anlysis. *Neural Networks*, 16(3-4):389–403.

[Webb, 2003] Webb, A. (2003). *Introduction to biomedical imaging*. Wiley.

[Weinstein et al., 1999] Weinstein, D., Strano, S., Cohen, P., Fields, S., Gomori, J. M., and Degani, H. (1999). Breast fibroadenoma: Mapping of pathophysiologic features with three-time-point, contrast-enhanced MR imaging – pilot study. *Radiology*, 210:233–240.

[Wismüller et al., 2002] Wismüller, A., Lange, O., Dersch, D., Leinsinger, G., Hahn, K., Ptz, B., and Auer, D. (2002). Cluster analysis of biomedical image time series. *International Journal of Computer Vision*, 46(2):103–128.

[Woods et al., 2004] Woods, B., Clymer, B., Saltz, J., and Kurc, T. (2004). A parallel implementation of 4-dimensional Haralick texture analysis for disk-resident image datasets. In *Proceedings of the ACM/IEEE SC2004 Conference*.

[Wu et al., 2004] Wu, T. F., Lin, C. J., and Weng, R. C. (2004). Probability estimates for multi-class classification by pairwise coupling. *Journal of Machine Learning*, 5:975–1004.

[Yoo et al., 2002] Yoo, S. S., Choi, B. G., Han, J. Y., and Kim, H. H. (2002). Independent component analysis for the examination of dynamic contrast-enhanced breast magnetic resonance imaging data. *Investigative Radiology*, 37(12):647–654.

[Zadrozny and Elkan, 2001] Zadrozny, B. and Elkan, C. (2001). Obtaining calibrated probability estimates from decision trees and naive Bayesian classifiers. In *Proceedings of the 18th International Conference on Machine Learning*, pages 609–616. Morgan Kaufmann, San Francisco, CA.

[Zou, 2002] Zou, K. H. (2002). Receiver operating characteristic (ROC) literature research. Online bibliography available from http://splweb.bwh.harvard.edu:8000/pages/ppl/zou/roc.html.