UNIVERSITÄT ZU LÜBECK

Course of Studies: Computational Life Science

Master Thesis

Automatic Classification of Morphological Patterns in Lung Tumor Tissue

Lisa Senger

Submission Date: 16. November 2011

Supervisor:	Advisor:	Medical Advisor:
Prof. Dr. Bernd Fischer	**André Homeyer**	**Dr. Frederick Klauschen**
MIC Institute of Mathematics and Image Computing	Fraunhofer MEVIS	CHARITÉ

Declaration of Authorship

I certify that the work presented here is, to the best of my knowledge and belief, original and the result of my own investigations, except as acknowledged.
Furthermore it has not been submitted, either in part or whole, for a degree at this or any other University.

Date *Signature*

Contents

1 Chapter 1 Introduction

The classification of different tumor types on microscopic images plays an important role in the diagnosis and therapy of tumors. In this work methods are presented to automate the classification of lung tumor tissue.

In clinical practice tissue sections of tumors are examined by a pathologist. Different tissue structures have to be identified and quantized during this step. The interpretation of the tissue depends on the knowledge and skills of the pathologist and errors could cause severe consequences for the patient.
Nowadays, the tissue sections can be digitalized which makes them available for computer aided analyzing methods. This opens the way for more objective interpretation of the tissue, which may help pathologists in their daily practice.

In this work existing tools that are used for other morphologic classification tasks are evaluated for their suitability in the classification of lung adenocarcinoma. For this lung tumor type new classification guidelines were recently presented in [23]. To classify lung adenocarcinoma automatically with respect to this guidelines, new analysis methods are needed to characterize the different tissue structures. Methods based on the discrete wavelet transform seemed promising for this task and are therefore tested and evaluated during this work.

The first goal in this thesis is to achieve classification results with a high detection rate of healthy lung tissue in order that it can be distinguished from tumor tissue. Furthermore the lung tumor tissue should be subclassified with respect to the new classification guidelines of lung adenocarcinoma. The focus lies on the evaluation of image attributes that characterize important structures of the adenocarcinoma subtypes. It is evaluated if it is possible to improve the classification with additional analyzing methods based on the discrete wavelet transform.

Outline of the thesis

The examination of microscopic tissue sections and the classification guidelines of lung tumor tissue are explained in more details in chapter 2. Furthermore, in this chapter the database of histological images and its characteristics along with difficulties in the classification are outlined.

The following chapters describe the methods used in the lung tumor classification. Chapter 3 introduces some general aspects of pattern recognition and the learning algorithm and software used for the classification are described.

The attributes of the histological images that are used as basis for the classification are introduced in the next chapters. Chapter 4 covers the image attributes that were already available for other classification tasks in microscopic images. One is based on the intensity values of the images, the other on local binary patterns. In chapter 5 on the other hand an additional image attribute based on the discrete wavelet transform is introduced.

The attributes are compared and evaluated in chapter 6. Furthermore results of the automatic classification of microscopic images of lung tumor tissue are presented. It is evaluated if a classification with methods based on the discrete wavelet transform improve the results.

In the last chapter a conclusion and an outlook to possible further research concerning the lung tumor classification is given.

2 Chapter 2 Medical Background

2.1 Lung Tumor

Lung carcinoma is the third most common cancer in Germany. The survival rate five years after the diagnosis lies between 13 % and 17 % for men and 13-19% for women [1]. For a better understanding of its pathogenesis and improvement of treatment options lung cancer is classified according to its molecular and morphologic variations. One type of lung cancer of the bronchus is discussed in more detail in this work, the so-called adenocarcinoma, a tumor that is derived from bronchial glands. It belongs besides the squamous cell carcinoma and the large cell carcinoma to the 'non-small cell lung cancer' (NSCLC). On the other hand there is the small cell lung carcinoma (SCLC). Currently the main treatment differentiation with respect to the histological type is based on the classification in NSCLC and SCLC. However, the conventional chemotherapy was refined in the last years and novel molecular markers were investigated, such as, for instance, activating EGFR (epidermal-growth-factor-receptor) mutations, which are related to the unregulated cell growing. Therefore ways are searched to correlate different histological subtypes according to their response to therapy or the presence of certain mutations.

The classification into one of the lung cancer subtypes is primarily done by histological examination of a biopsy taken of tumor tissue or of an excised tumor. Sometimes a radiological picture can give a first indication of the tumor type but appropriate treatment decisions can only be made after the histological examination [10]. The acquisition of

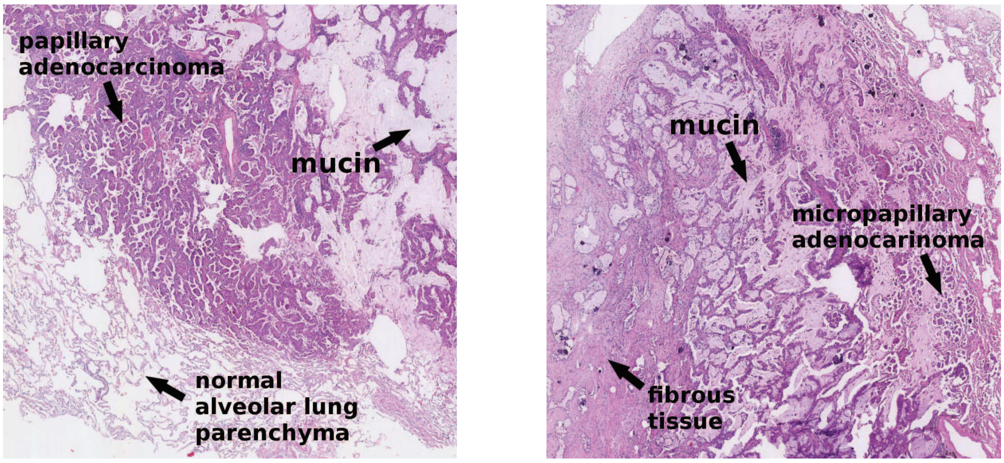


Figure 2.1: Parts of histological tissue sections of lung adenocarcinoma.

histological tissue is explained later in section 2.2.1.

The subclassification of the invasive adenocarcinoma is investigated in more detail during this work. The adenocarcinoma is the most common histological lung cancer subtype [23]. It consists of tumor cells emerged from gland tissue and is mostly localized in the peripheral lung [10]. Microscopic images of lung adenocarcinomas can be found in figure 2.1.
With respect to the enhanced tumor therapy and investigation of novel molecular markers, the old classification of adenocarcinoma is insufficient. Therefore a new classification guideline to determine the different subtypes of invasive adenocarcinoma is introduced in [23]. As a result the adenocarcinoma should be divided into the following five histological subtypes: acinar, papillary, micropapillary, lepidic and solid adenocarcinoma. Before the main characteristics of the subtypes are presented, the characteristics of healthy lung tissue are described now.

The major tissue components in the lung are alveoli, bronchi and blood vessels. Moreover, there is cartilage, connective tissue, nerves, serous and mucous glands. The inhaled air streams through the trachea, passes the bronchi, which branch further into the bronchiole until the alveoli are reached. Here the gas exchange takes place. In figure 2.2 two microscopy images of the lung are shown. Alveoli compose the largest lung

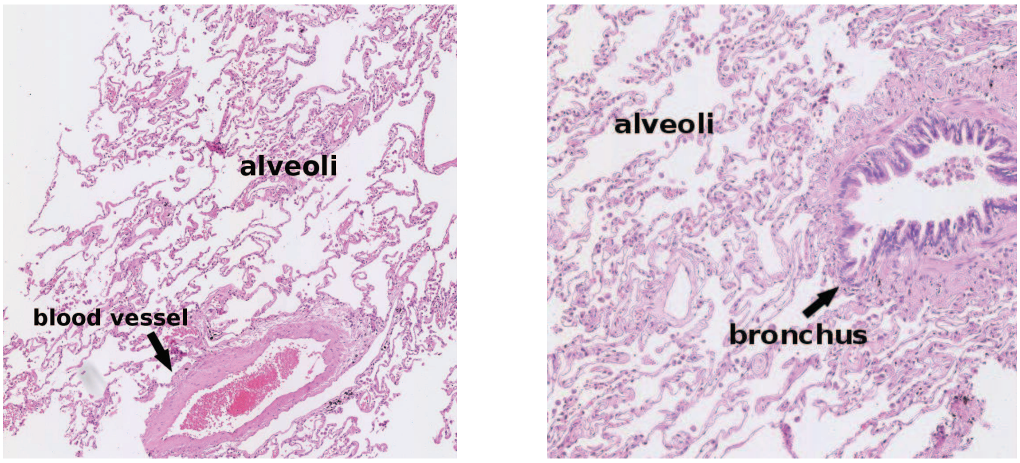


Figure 2.2: Parts of histological sections of normal lung tissue.

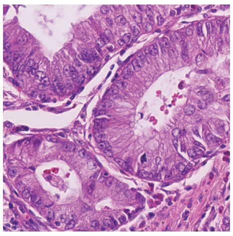
tissue component. They are separated by thin walls which contain the capillaries. In the left picture of figure 2.2 also a blood vessel is visible, in the right one a bronchus. In comparison with figure 2.1 one can clearly see the differences to tumor tissue, which has lost all air filled spaces.

Microscopy image parts of the adenocarcinoma subtypes can be seen on the left hand side of figure 2.3. The main characteristics are described on the right hand side (sources of the descriptions:[23, 10]). An invasive adenocarcinoma can consist of one of these histological subtypes and also of mixtures of the subtypes. Therefore the tumor is classified according to the predominant subtype. Additionally, variations of the subtypes may occur, for example with a production of mucin, like in the sections on figure 2.1. In [23] a classification based on these subtypes is recommended. As a result non-mucious lepidic adenocarcinoma has the best, the solid and micropapillary subtypes the worst and papillary and acinar adenocarcinoma have an intermediate prognosis.

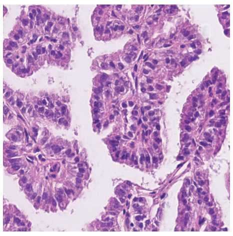
2.2 Database: Histological Images

2.2.1 Data Acquisition: Tissue Sample Processing

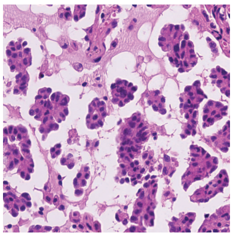
A tissue sample, taken for example from a biopsy, need to be processed in several steps until it is ready for a microscopic examination. These steps are explained now.



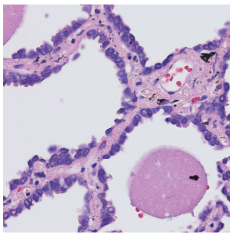
acinar adenocarcinoma
Consists of round or oval shaped glands with a central lumina.



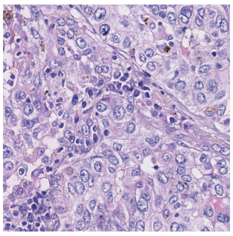
papillary adenocarcinoma
Consists of glandular cells growing along central fibrovascular cores.



micropapillary adenocarcinoma
Consists of papillary nodes, which contain in contrast to papillary adenocarcinoma no fibrovascular cores.



lepidic adenocarcinoma
Appears well differentiated. The tumor cells are spread along the surface of alveolar walls, without destroying them.



solid adenocarcinoma
Appears undifferentiated. The polygonal tumor cells are arranged in compact formations. No acinar, papillary, micropapillary or lepidic growth is recognizable.

Figure 2.3: Histological subtypes of invasive adenocarcinoma.

Fixation

The aim of fixation is that the tissue and cells change their natural appearance as little as possible. Therefore the tissue is put in most cases into a formalin solution shortly after the removal, in which it is chemically fixated [15].

Embedding

Before thin tissue section can be cut, the consistency of the sample needs to become firmer. Therefore, the tissue is first dehydrated, for example by the substitution of water with ethanol [15]. Afterwards the tissue is embedded for hardening for example in a paraffin solution [15].

Sectioning

After hardening the tissue sample block can be cut into small slides with a microtome [15]. The standard thickness of a section lies between 5 and 8 μm [11]. The sections are placed on a microscope slide for further processing.

Staining

Staining is used to enhance contrast in the microscopic image by highlighting important structures. Depended on the technique different parts of the tissue are highlighted. The histological sections in this work are stained with the H.E.-method. The two used stains are hämatoxylin and eosin. Hämatoxylin colors the cell nuclei blue, while eosin colors the cytoplasm and collagen in red or pink.

For this work several images of tissue sections of lung adenocarcinoma were provided by Dr. Frederick Klauschen, Charité Berlin. They were digitalized with the Hamamatsu NanoZoomer. The scanned sections are available on different magnifications with a highest resolution of 0.23 μm/pixel. The size of a scanned section on this magnification is on average approximately 200.000 × 100.00 pixel.

All in all 22 images of lung tumor tissue sections were available for the classification in this work. Image regions of 1024 × 1024 pixel were classified into one of the adenocarcinoma subtypes or into a healthy tissue class. This corresponds to a tissue area just under 250 × 250 μm. The image regions in figure 2.3, 2.4 and 2.5 are of this size.

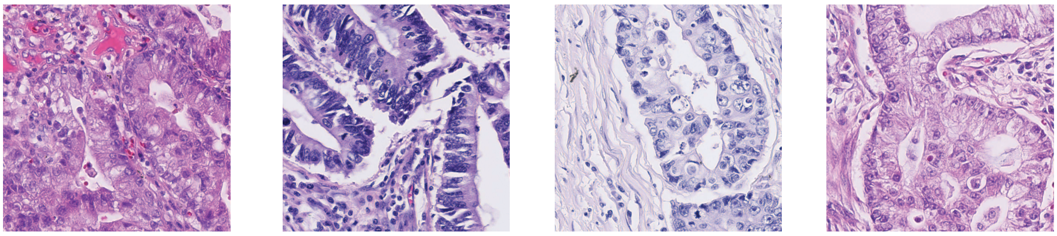
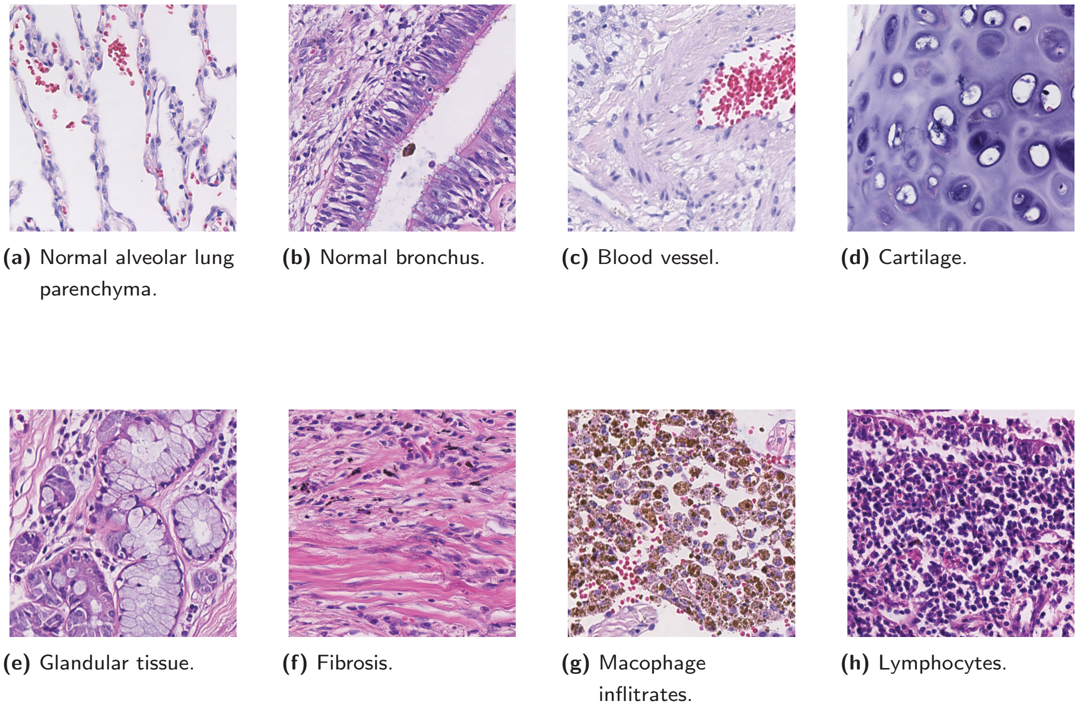


Figure 2.4: Inner-class differences of acinar lung adenocarcinoma.

2.2.2 Database Characteristics

The aim in this work is to classify and quantify the different lung adenocarcinoma subtypes in a microscopic tissue section. The main characteristics of the subtypes were already explained in section 2.1. Now further properties and characteristics of microscopic images of tumorous lung tissue are discussed.
One of the difficulties in the typification of lung adenocarcinoma is that the appearance of one subtype often varies. Inner-class differences are for example produced by variations in the staining intensity, which can be seen in 2.4. This can occur due to variations during the preparation procedure, slightly different section thickness or physiological variations from different patients. But also the tissue structure itself can vary within one class. This can also be seen in figure 2.4 in which the structures differ from each other. These example subimages however can still clearly be assigned to the acinar adenocarcinoma subtype. But this is not always the case. In some cases an exact classification is even for a specialist difficult to accomplish.
Furthermore the adenocarcinoma subtypes have to be distinguished from other tissue types that can occur in lung tissue. Some of these are shown in figure 2.5. Parts of biopsies of lung tumors also include healthy tissue. This can be lung specific tissue like alveoli or bronchi, as well as other healthy tissue like blood vessels, cartilage or glandular tissue. Often, there are also immune reactions in combination with a tumor. This goes along with the appearance of lymphocytes, macrophages or fibrosis. All of this tissue types and cells have to be recognized as well as the adenocarcinoma subtypes to distinguish between healthy and tumorous tissue and are therefore also considered in the classification task in this work.
All in all 14 tissue classes are distinguished in this work, containing the five adenocar-



(a) Normal alveolar lung parenchyma.

(b) Normal bronchus.

(c) Blood vessel.

(d) Cartilage.

(e) Glandular tissue.

(f) Fibrosis.

(g) Macophage inflitrates.

(h) Lymphocytes.

Figure 2.5: Histological images of healthy tissue inside the lung. Besides lung specific tissue, like aveoli or brochni, also other tissues have to be distinguished from the adenocarinoma types.

cinoma subtypes of figure 2.3, mucious adenocarcinoma and the healthy tissue classes shown in figure 2.5.

Chapter 3

3 Pattern Recognition in Histological Images

Like explained in the previous chapter the discrimination of different tissue classes in histological images is not an easy task and in many cases the types are not clearly distinguishable. This is due to the great variability of the tissue, even if it has the same type. A pathologist needs a lot of experience to give reliable results. The same holds, if the tissue should be classified automatically. If one want to create an algorithm for this classification task, it should somehow gain experience, like the pathologist, by training on many examples. This procedure is called *machine learning*. The main aspects are explained now, more details can for example found in [24].

The goal of machine learning is to solve a given problem with example data or past experience using an algorithm. Pattern Recognition in particular deals with the classification of data into several groups. This can be done using example data for which the group of each example, the ***class*** or ***label***, is already known. In a first step an algorithm is trained with this data. It determines which characteristics or values of the several attributes of the examples, the so called ***features***, are similar given a particular class. Afterwards the trained algorithm can be used to determine the classes of a second data set, which is not labeled. This method is called ***supervised learning***. In figure 3.1 an example of a data set with two classes is visualized. Possible discrimination curves of two classifiers for this data are shown. In this two dimensional example a new data

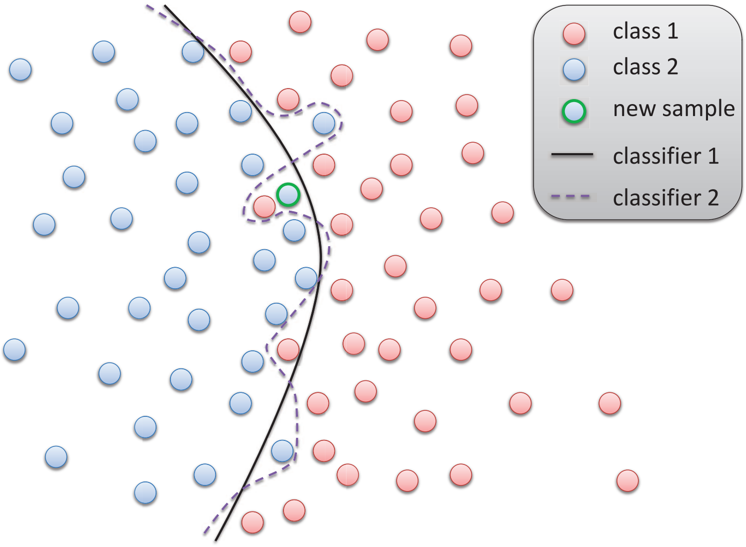


Figure 3.1: Example classification results of two classifiers.

point is classified depending on which side of the discrimination curve it lays. In the image this new sample is denoted with a green border. The true underlying class is the blue one. With classifier 1 the sample is correct classified, with classifier 2 it is not. Example data can also be classified using an ***unsupervised learning*** algorithm. In this case the example data is not yet grouped into different classes, but unlabeled. The algorithm defines by itself, which examples have similar characteristics and are therefore identified as one class.

The classification of the lung tumor classes in this work is done with a supervised learning algorithm. Before the algorithm is trained some aspects have to be considered. The training data set should be a representative set of examples, which covers as many variations of one class as possible. Secondly, the learning algorithm and its parameters should be chosen in a way that it does not overfit the data. Overfitting means that the algorithm can classify the training data set very well, but fails if it is applied to new data. This often happens when the algorithm is trained too much. In figure 3.1 an example of overfitting can be seen. Classifier 2 perfectly discriminates the training data but classifies the new sample wrong. Classifier 1 on the other hand has errors on the

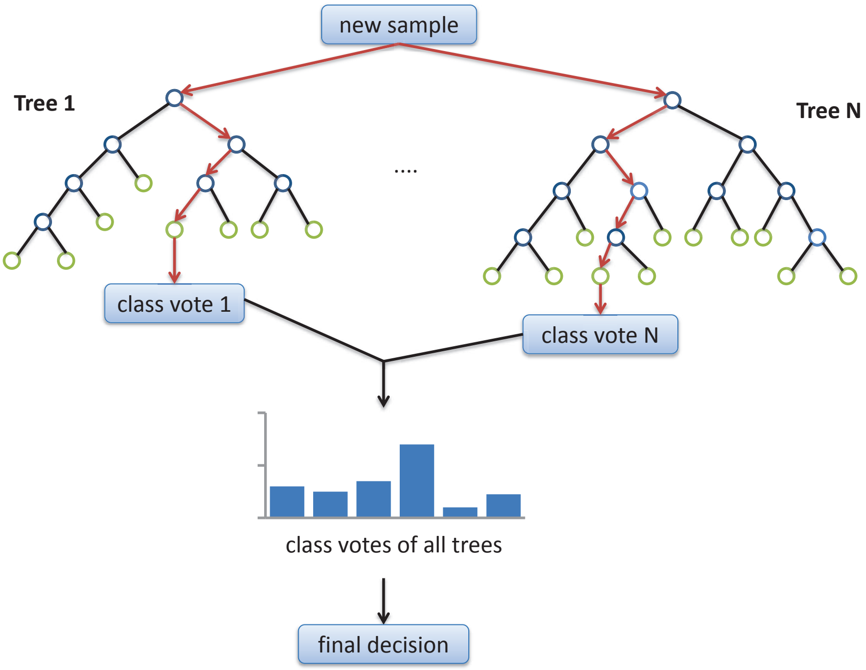


Figure 3.2: Schematic illustration of the classification of a new sample with random forests.

training data, but the new sample is nonetheless classified correct. This is one of the main challenges in pattern recognition: The algorithm should perform as well as possible on the training data *and* new data.

3.1 Random Forest Classifier

The random forests learning algorithm, introduced by L. Breiman [2], is used in this work to classify the different lung tumor tissues. It consists of many decision trees, which assign the most popular class to an input vector.
In a decision tree the leave nodes represent class labels and at the branch nodes conditions are set to the features that lead to this labels. A new sample is routed down the tree according to the values of its features. It is assigned to the class it reaches at the leaf node. There are several methods to build a decision tree. Some are for example given

in [24].

A single decision tree is fast but also tends to overfit the data. With random forests this weakness is disposed by using many decision trees on randomized subsets of the training data. The generation of random forests is explained now.
Let's assume we have a training data set with N examples and M features. Each tree in the forest is built using a new training set of size N, where the examples are selected at random with replacement form the original training data set. At each node $m << M$ features are chosen randomly and used to find the best split. Each tree is fully grown and not pruned. For a new sample a label is assigned with each of these trees. This is illustrated in figure 3.2. The final prediction of the random forest is the class with the most votes of all trees.

According to [2] the error rate of a random forest depends on two things. First, the correlation between the trees in the forest. If the correlation gets higher, the error rate increases. And secondly the error rate of each individual tree. If it is kept low, the error rate of the random forest is also decreasing. These two points are adjustable with the choice of m. If m is reduced, the correlation between the trees is reduced but at the same time the error rate of the individual trees is increased, and vice versa if m is increased. So the goal is to find a compromise, where the final error rate is 'optimal'. According to [2] $m = int(log_2M + 1)$ is a good choice and therefore used in this work. The number of trees used is 25.

3.2 Classification Software: HistoCAD

The classification of histological lung tumor tissue is done with the application HistoCAD. It was developed by André Homeyer at Fraunhofer MeVis. With HistoCAD histological images can be analyzed and classified according to the tissue characteristics that are relevant for a certain diagnostic task.
A screenshot of the software is shown in figure 3.3. On the left hand side an image of a microscopic tissue section is displayed. This image is already preprocessed automatically. During preprocessing the image is overlaid with a grid from which tiles with no tissue are removed. This segmentation is based on the intensity values of the tiles.

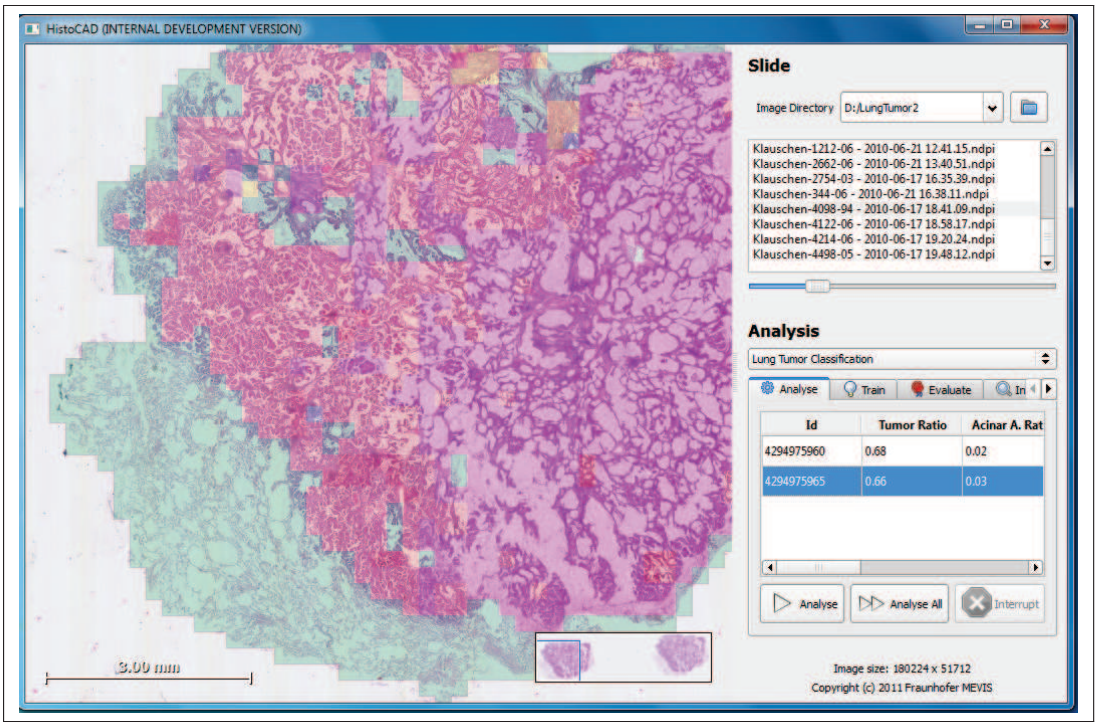


Image Directory: D:/LungTumor2

Klauschen-1212-06 - 2010-06-21 12.41.15.ndpi
Klauschen-2662-06 - 2010-06-21 13.40.51.ndpi
Klauschen-2754-03 - 2010-06-17 16.35.39.ndpi
Klauschen-344-06 - 2010-06-21 16.38.11.ndpi
Klauschen-4098-94 - 2010-06-17 18.41.09.ndpi
Klauschen-4122-06 - 2010-06-17 18.58.17.ndpi
Klauschen-4214-06 - 2010-06-17 19.20.24.ndpi
Klauschen-4498-05 - 2010-06-17 19.48.12.ndpi

Lung Tumor Classification

Id	Tumor Ratio	Acinar A. Rat
4294975960	0.68	0.02
4294975965	0.66	0.03

Image size: 180224 x 51712
Copyright (c) 2011 Fraunhofer MEVIS

Figure 3.3: Screenshot of HistoCAD.

If a tile contains too little color information is it declared as background and removed. For the lung tumor classification an expanded segmentation is used. Additionally tiles are kept whose neighbors achieve the necessary intensity values in order to reduce segmentation gaps in alveolar tissue. The tile size in the lung tumor classification is 1024 × 1024 pixel.

With the panel shown on the right hand side of figure 3.3 the user can select images and start the classification of their tiles. Therefore the images have to be 'analyzed', which means that for every tile of the images different features are calculated. The available features are based on intensity values and local binary patterns. For the lung tumor classification wavelet based features were implemented additionally. These image attributes are explained in the next sections. All features can be calculated on different magnifications that are available with histological images and on different color channels. In order to keep the software complexity low the features are selected by the developer with respect to a certain diagnostic task and cannot be changed by the user.

After the images are analyzed a learning algorithm can be trained. The training data can be built by the user by selecting tiles and refer them to a class. In this way also an existing training set can be extended, for example by adding falsely classified tiles with their true class to the training set. For simplicity of the application the appropriate learning algorithm is also chosen by the developer. In the lung tumor classification random forests are used to classify the tiles. Other available learning algorithms are Nearest Neighbor and Naive Bayes, see [24] for more information.

The classification result is visualized in different colors for each tissue class. For example the lung tumor tissue section in figure 3.3 has mainly been classified as normal alveolar lung parenchyma (green), papillary adenocarcinoma (orange) and mucious adenocarcinoma (pink). In the lung tumor classification the HistoCAD software additionally displays the ratios of the five lung adenocarcinomas as well as the overall tumor ratio after the classification.

Chapter 4

4 Extraction of Characteristic Features in Histological Images

In this chapter the features used for the lung tumor classification are described. First a short introduction to feature extraction in histological images is given. In the sections 4.2 and 4.3 the features are described which were already available for other classification tasks in HistoCAD. The new feature, which was implemented specially for the lung tumor classification in this work, is explained in the next chapter. The last part of this chapter deals with feature selection methods.

4.1 Introduction to Texture Features

The aim of this work is to characterize histological tissue samples automatically with a classification algorithm. For this the tissue section is overlaid with a grid and each tile is classified separately, like explained in chapter 3.2. The classifier was already described, now the focus lies on the feature vector which is used as basis for the classification.

Theoretically the image tile can be classified using all image pixels. However, this would result in a huge input data set, which contains a lot of redundant information. The better way is to extract special features from the image, which describe the main characteristics more precisely with less values. These extracted features are combined to a feature vector. The chosen features should be discriminating and as independent as possible to achieve good classification results. But besides this the calculation time should not be disregarded. Working with histological images often means working with huge data sets, as explained in chapter 2. It is not unlikely that one image of a tissue section has a size of 200 000 x 100 000 pixels. With a tile size of 1024x1024 pixel this results in approximately 20 000 image regions. This means that the texture features have to be calculated a lot of times and this in the best case in a few minutes. It follows that the calculation of the texture features should not be too complex. A feature that improves the classification result is practically only valuable if the calculation can be done in a short time.

Dependent on the task, there are many possible attributes of the image that can be used as feature vector. For example one can extract certain properties of the image, like edges, the luminance or the occurrence of a particular color value. In this work the images are histological, which makes the texture an important attribute. In the literature many different texture features can be found. For example simple statistical texture features like color histograms and co-occurrence matrices. Often used are also Garbor filters and features based on the discrete wavelet transform, where it is tried to capture the image structures in a similar way like in the model of the human visual system.
Especially texture features based on wavelets are very common. With them high classification accuracies can be achieved, whereas their complexity, for example in comparison to Garbor filters or co-occurrence matrices, is low [20]. Especially in the analysis of histological images they are often used and show are great promise for good classification results for this kind of images [12, 13].
For this reason wavelet based features are evaluated for their suitability as an additional texture feature for the lung tumor classification part of the HistoCAD application. All in all three different kinds of features are used for the classification, where two were already available. The first comprises statistics of the pixel values and the second is based on local binary patterns. They are described in the next two sections. The texture feature based on wavelets is described in the next chapter.

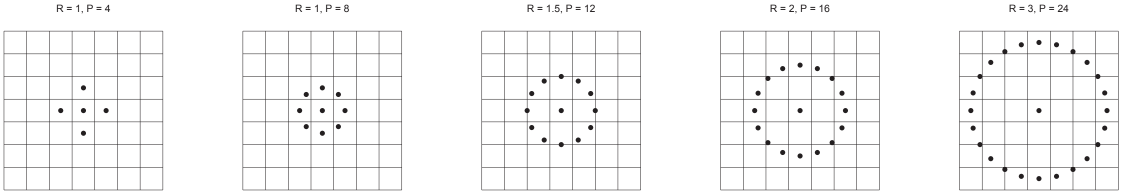


Figure 4.1: Different examples of the pixels used for the calculation of a local binary pattern. The P neighbor pixels lie on a circle with radius R around the center pixel.

4.2 Intensity Attributes

The most simple texture features in HistoCAD are based on the intensity statistics of the pixels in an image region. At the same time they are already very discriminative features, which can be seen later in chapter 6. For a predefined image region the following intensity statistics are calculated and used as texture feature: minimum value, maximum value, sum of all values, mean, standard deviation, lower quartile, median and upper quartile.

The discriminative power of the simple image statistics can be explained with the special characteristics of microscopic images. The histological stains highlight different structures in the tissue with different colors. This means a lot of information about the tissue is given by its color. Therefore a lot of information is also revealed by the intensity statistics of the image.

4.3 Local Binary Patterns

One of the texture features used in this work to classify the histological images are local binary patterns, which are described in [19]. These texture features are theoretically very simple and effective at the same time. Advantages are that the local binary patterns are gray scale and rotation invariant and can be calculated on different resolutions. The main idea is to determine a pattern from the circular neighborhood of a pixel in an image region. The occurrences of different patterns in this region are stored in a histogram which gives the feature vector.

A local binary pattern of an image region it calculated with a center pixel g_c and P

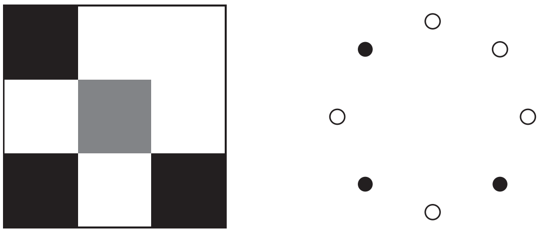


Figure 4.2: A possible 3×3 image region on the left hand side and its resulting local binary pattern on the right. A black dot refers to the binary number 0 and a white one to 1.

neighboring pixels $g_0, ..., g_{P-1}$ on a circle with radius R around the center. With different choices of P and R, one can adapt the resolution of the pattern to particular characteristics of the image texture. This is shown in figure 4.1. Points that do not directly lie in the center of a pixel are determined with interpolation. In this work an image region of size 3×3 is used as a pattern, which corresponds to $P = 8$ and $R = 1$, see the second image in figure 4.1.

With these pixels the local binary pattern is calculated as follows. First the center pixel is subtracted from each of the circularly distributed neighbor pixels. Afterwards it is determined whether the neighbor pixels $g_p (p \in \{0, ..., P-1\})$ have smaller or greater gray values than the center pixel g_c. The result is encoded in a binary number where 1 corresponds to a gray value greater than the center value and 0 to a smaller value. An additional factor 2^p encodes the position. An example is given in figure 4.2. A 3×3 example image region is on the left hand side, the corresponding binary pattern 11101010_2 on the right side. The whole spatial structure of the image region is characterized by the sum of the weighted binary number or in other words by converting the binary number to a decimal number. In this example it holds that $LBP_{8,1}('example\ pattern') = 87$.

Mathematically a local binary pattern of an image region is described with:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p,$$

with

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}.$$

The pattern is gray scale invariant because the difference taken in the first step is not affected by changes in the luminance. However, the pattern is not yet rotation invariant.

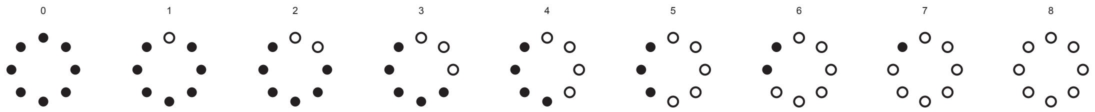


Figure 4.3: The 9 'uniform' local binary patterns.

This is due to the factor 2^p which makes it possible to detect the exact position of the binary values in the pattern.

Rotation invariance is achieved by detecting only 'uniform' patterns. These are patterns have only a limited number of transitions from 0 to 1 or vice versa. Due to [19] these uniform patterns are with over 90 % the most common in examined surface structures. For this a uniformity measure $U('pattern')$ is introduced, which detects the number of 0/1 transitions in the pattern. A local binary pattern is called uniform if the number of bitwise 0/1 changes is smaller or equal 2. This results in 9 different uniform patterns, shown in figure 4.3. Each of these patterns detects a particular image structure. For example the uniform patterns #0 and #8 detect dark and bright spots respectively, whereas #4 detects edges.
The operator for gray-scale and rotation invariant texture description is given by

$$LBP_{P,R}^{riu2} = \begin{cases} \sum\limits_{p=0}^{P-1} s(g_p - g_c), & \text{if } U(LBP_{P,R}) \leq 2 \\ P+1, & \text{otherwise,} \end{cases}$$

where

$$U(LBP_{P,R}) = |s(g_{P-1} - g_c) - s(g_0 - g_c)| + \sum_{P=1}^{P-1} \left| s(g_p - g_c) - s(g_{p-1} - g_c) \right|.$$

The superscript ^{riu2} describes that rotation invariant uniform patterns are used, where the uniform patterns contain lower or equal to 2 bitwise 0/1 changes.

The final local binary texture feature of an image is determined with an occurrence histogram of the 10 different local binary patterns $LBP_{P,R}^{riu2}$ in the image. An evaluation of this texture feature is given in chapter 6.

4.4 Feature Selection

In the lung tumor classification task in HistoCAD the described features can be calculated on the different magnifications that are available with every histological image and on different color channels. Calculating all possible feature goes along with an immense increase in image-analysis time and is therefore not practicable. Thus, a method is needed to select just as many features to solve the classification task in an appropriate way whereas the calculation time is kept low. Removing irrelevant and redundant information from the feature set may additionally allow the classification algorithm to operate faster and more effectively. In some cases feature selection also improves the classification accuracy.

The first step in the selection of an appropriate feature subset is the definition of a measure that determines the quality of the subset. There are several approaches for this measure. For example one can determine the performance of the classifier for the subsets and choose the one with the best. In this work a measure is used which determines the dependencies of the feature to each other and the classes. It is described in section 4.4.1.

After a validation method for a feature subset is selected, the best subset has to be found. Therefore the space of all possible subsets has to be searched. In most cases an exhaustive search is impracticable because the feature set is too big. Therefore the space is typically searched greedily. One greedy feature selection method is the forward selection. It starts with zero features and adds new features in each step until the addition does not improve the quality measure. A similar method starts with the full feature set and reduces it in each step. It stops if the elimination of features does not improve the selection. This method is called backward elimination. More feature selection methods can be found for example in [24]. In this work the feature set is selected with forward selection in combination with a correlation based subset measure which is explained now.

4.4.1 Correlation-based Feature Selection

M. Hall developed in [8] a correlation-based feature selector (CFS). The goal was to create a feature selection method that eliminates redundant features whereas features that are highly predictive of the class are kept. He states:

> *"A good feature subset is one that contains features highly correlated with (predictive of) the class, yet uncorrelated (not predictive of) each other."* [8, p.52]

Based on this thesis he built a measure M_S to evaluate a feature subset S:

$$M_S = \frac{k\overline{r_{cf}}}{\sqrt{k+k(k-1)\overline{r_{ff}}}},$$

where k is the number of features in S, $\overline{r_{cf}}$ the mean class-feature correlation ($f \in S$) and $\overline{r_{ff}}$ the feature-feature inter-correlation. To calculate M_S the training data is first discretized before the class-feature and feature-feature correlations are determined with symmetric uncertainty. With one of the greedy search strategies explained above the feature subset with the highest M_S is found.

In this work the feature selection is done with Rapid Miner [18]. Rapid Miner, formerly named YALE (Yet Another Learning Environment) is a data mining software that provides a lot of methods for machine learning procedures. Amongst others it contains methods for data preprocessing, visualization, modeling and evaluation. Furthermore the machine learning algorithms of WEKA [9] can be included. This extension also contains an implementation of correlation based feature selection.
With this tools a work flow is built in this work to select an appropriate feature subset using forward selection in combination with CFS. The results can be found in chapter 6.

5 Chapter 5

Wavelets and Their Application in Feature Extraction

In this chapter an additional texture feature is presented which was implemented for the texture classification in the histological approach. It uses the information in the image given by the discrete wavelet transform. Wavelet based features are often used in texture classification, for example in [12, 13, 4, 16]. Research on texture analysis has shown that with methods based on wavelets high classification accuracies can be achieved [21]. Because the discrete wavelet transform is based on multiresolution, it gives the opportunity to asses scale dependent information of the texture. This is especially useful in the analysis of complex structures, like histological images. Here the differences of images are often only recognizable if they are compared on different resolutions. Furthermore, in [12] a wavelet based feature is proposed which detects anisotropic structures in an image. This gives the possibility to distinguish certain tissue structures, like fibrosis, more clearly from other tissues that are not oriented in a preferred direction.

This chapter is structured as follows. First an introduction to the theory of the discrete wavelet transform is given. This description is at some points kept short and not every detail is explained, only the ones needed to understand the interpretation of the transform and its implementation. Deeper information about the discrete wavelet transform or wavelets in general can be found in the literature, for example in [7, 3, 22].

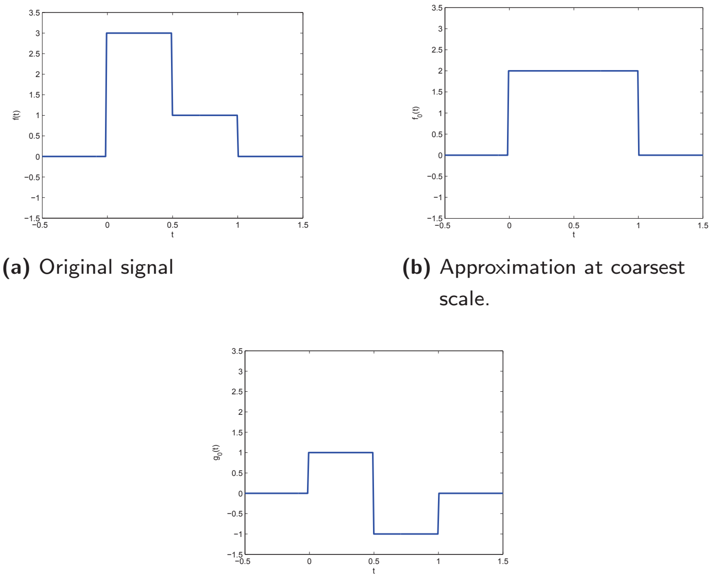
In section 5.2 the filterbank implementation of the discrete wavelet transform is described. It requires a basic understanding of filters and convolution, which is not discussed here. Afterwards, influences of the choice of different wavelet bases are discussed in chapter 5.3. The calculation of wavelet based features and their interpretation is described in the last part of this chapter.

5.1 Introduction to the Discrete Wavelet Transform

If one sees an image, the human visual system first recognizes the main information of the picture. Details are not captured with the first look but one can already determine the main aspects of the image. Only with longer observation one sees more and more details until the whole image is recognized.
With the discrete wavelet transform it is possible to transfer this idea to the analysis of functions. A function f is first captured at a coarse scale and then the details are determined by going on finer and finer scales. To derive the calculation of the discrete wavelet transform f is described on different scales. For this vector spaces are needed which correspond to these scales. The union of these spaces gives the 'whole space' in which the original function f lives. In the continuous case this 'whole space' is generally given by the Hilbert space $L^2(\mathbb{R})$. This approach is called multiresolution. The usage of multiresolution in the context of the discrete wavelet transform was first published by Mallat and Meyer in 1988/99, see e.g. [17], and the main aspects are summarized now.

Let's assume that we want to represent a signal $f(t) \in L^2(\mathbb{R})$ at different scales. For the most purposes it is sufficient to consider only resolutions along the dyadic sequence $(2^j)_{j\in\mathbb{Z}}$. In this work, the coarsest scale corresponds to $j = 0$. With increasing j, the scale gets finer. In the discrete case this means that the number of discretization points double whereas their distance halves.
For the explanation of multiresolution let's take a simple example signal $f(t)$, given by

$$f(t) = \begin{cases} 3, & 0 \leq t < \frac{1}{2} \\ 1, & \frac{1}{2} \leq t \leq 1 \, . \\ 0, & \text{otherwise} \end{cases}$$



(a) Original signal

(b) Approximation at coarsest scale.

(c) Details.

Figure 5.1: Example signal and its decomposition with the Haar wavelet.

The signal is shown in figure 5.1a. This signal is first approximated at the coarsest scale. This means that $f(t) \in L^2(\mathbb{R})$ is projected onto another vector space $\mathcal{V}_0$.
For the approximation a set of basis functions of $\mathcal{V}_0$ is needed. Therefore we need the so-called ***scaling function*** $\phi(t) \in L^2(\mathbb{R})$. $\mathcal{V}_0$ is then defined as the space with all combinations of $\phi(t)$ and its shifts $\phi(t-k), k \in \mathbb{Z}$. We can choose for example

$$\phi(t) = \begin{cases} 1, & 0 \leq t \leq 1 \\ 0, & \text{otherwise} \end{cases}.$$

This scaling function is called Haar-scaling function. With this $\phi(t)$, $\mathcal{V}_0$ contains all functions that are constant on intervals of length 1. The original signal $f(t)$ can now be approximated at the coarse scale with $j = 0$ with a linear combination of the basis functions:

$$f_0(t) = \sum_k a_0(k)\phi(t-k) \quad ,(k \in \mathbb{Z}),$$

where $f_0(t) \in \mathcal{V}_0$ is the approximation of f at resolution 2^0 and $a_0(k) \in \mathbb{R}$.
The best approximation of the example signal at resolution 2^0 is given if the Haar basis

function is multiplied with 2. It holds that

$$f_0(t) = 2 \cdot \phi(t) = \begin{cases} 2, & 0 \leq t \leq 1 \\ 0, & \text{otherwise} \end{cases},$$

which is shown in figure 5.1b.

To represent $f(t)$ at other resolutions 2^j, it is projected onto other vector spaces $\mathcal{V}_{2^j}$. The orthonormal bases of this spaces are build by additionally dilating $\phi(t)$. The dilations and translations of $\phi(t)$ are given by $\phi_{2^j,k}(t) = \sqrt{2^j}\phi(2^j t - k), k \in \mathbb{Z}$. It holds that there exists a $\phi(t) \in L^2(\mathbb{R})$ such that

$$\{\phi_{2^j,k}(t)\}_{k\in\mathbb{Z}} \text{ is an orthonormal basis of } \mathcal{V}_{2^j}. \quad [17] \tag{5.1}$$

That means $f(t)$ can be approximated at any resolution 2^j with:

$$f_{2^j}(t) = \sum_k a_j(k)\sqrt{2^j}\phi(2^j t - k) \quad ,(k \in \mathbb{Z}),$$

where $f_{2^j}(t) \in \mathcal{V}_{2^j}$ is the approximation of $f(t)$ at resolution 2^j. The coefficients $a_j(k)$ are given with $a_j(k) = \langle f, \phi_{2^j,k} \rangle$.

Note that the accuracy of the approximation is closely connected to the properties of ϕ. If ϕ is for example a rectangular function, like in our example, the approximation of a step function is very good but in contrary a smooth function cannot be approximated very well.

Lets now take a closer look at the vector spaces $\mathcal{V}_{2^j}$. They have to satisfy some properties to from a multiresolution approximation of $L^2(\mathbb{R})$, given in [17]. For example it is necessary that

$$\mathcal{V}_{2^j} \subset \mathcal{V}_{2^{j+1}}, \quad \forall j \in \mathbb{Z}$$

and

$$f(t) \in \mathcal{V}_{2^j} \Leftrightarrow f(2t) \in \mathcal{V}_{2^{j+1}}.$$

This means that the spaces at different resolutions are nested and that the elements in one space are scaled versions of the elements in the next space. It follows that if $\phi(t) \in \mathcal{V}_0$, it also holds that $\phi(t)$ is in $\mathcal{V}_1$. Because of (5.1), $\phi(t)$ can be expressed in the vector space $\mathcal{V}_1$ as a linear combination of $\phi_{2^1,n}(t)$:

$$\phi(t) = \sqrt{2}\sum_n h_0(n)\phi(2t-n), \quad n \in \mathbb{Z}. \tag{5.2}$$

This equation is called ***dilation equation*** or ***refinement equation*** [22]. The coefficients h_0 are a low-pass filter with $\sum_n h_0(n) = \sqrt{2}$ and are called ***scaling function coefficients*** [22]. They are needed to calculate the discrete wavelet transform of a signal, which will be described later.

But how to choose the scaling function $\phi(t)$ and how do we get the coefficients h_0? One can see that (5.2) is like an differential equation with coefficients h_0 and solution $\phi(t)$ that probably not even exists. So one can choose h_0 and try to solve the equation to determine the $\phi(t)$. But as described later this section, it is not necessary for the discrete wavelet transform to have a closed form for the scaling function $\phi(t)$. The goal is to choose h_0 in a way, that the scaling function has nice properties in order that it gives good a approximation of the original data. Ingrid Daubechies showed that it is possible to derive properties of the scaling function by properties of the scaling function coefficients [6].

The coefficients derived by Daubechies will be presented later in section 5.3. Now the role of the wavelet function in the discrete wavelet transform will be explained.

If the function $f(t)$ is approximated at a certain resolution, some information gets lost. To extract this lost information, the differences of the approximations at two successive resolutions 2^j and 2^{j+1} have to be known. For the example from the beginning the difference between the original and the approximated signal can easily be determined by looking at the plots of these two function in figure 5.1. The difference is shown in figure 5.1c.

To determine the difference mathematically, the orthogonal complement of $\mathcal{V}_0$ is needed [17]. Let $\mathcal{W}_0$ be this vector space for which holds:

$$\mathcal{V}_1 = \mathcal{V}_0 \oplus \mathcal{W}_0,$$

which means that $\forall f_1 \in \mathcal{V}_1 \; \exists f_0 \in \mathcal{V}_0, g_0 \in \mathcal{W}_0$: $f_1 = f_0 + g_0$. That means the information that is lost by going from $\mathcal{V}_1$ to the coarser resolution in $\mathcal{V}_0$ is represented by the vector space $\mathcal{W}_0$. In general the vector space $\mathcal{W}_{2^j}$ contains the details of the signal at resolution 2^j. It follows that the 'whole space' $L^2(\mathbb{R})$ can now be represented by combining the approximation in $\mathcal{V}_0$ with the details in $\mathcal{W}_{2^j}$, i.e.

$$L^2(\mathbb{R}) = \mathcal{V}_0 \oplus \mathcal{W}_0 \oplus \mathcal{W}_1 \oplus \mathcal{W}_2 \oplus ... \tag{5.3}$$

To calculate the details of $f(t)$ at resolution 2^j in the vector space $\mathcal{W}_{2^j}$, again a set of basis functions is needed. Let $\psi(t) \in L^2(\mathbb{R})$ be the so-called ***wavelet function***. The basis can be build with dilations and translations of ψ given by $\psi_{2^j,k}(t) = \sqrt{2^j}\psi 2^j t - k$. It holds that there exists a $\psi(t)$ such that

$$\{\psi_{2^j,k(t)}\}_{k\in\mathbb{Z}} \text{ is an orthonormal basis of } \mathcal{W}_{2^j}. \quad [17] \tag{5.4}$$

That means the details of $f(t)$ can be calculated at any resolution 2^j with:

$$g_{2^j}(t) = \sum_k d_j(k)\sqrt{2^j}\psi(2^j t - k) \quad ,(k \in \mathbb{Z}),$$

where $g_{2^j}(t) \in \mathcal{W}_{2^j}$ is the detail function of $f(t)$ that is lost by going from resolution 2^j to 2^{j+1}. The coefficients $d_j(k)$ are given with $d_j(k) = \langle f, \psi_{2^j,k} \rangle$.
Let's go back to our example and look again at the difference of the original function and its approximation in figure 5.1c. Here

$$g_0(t) = \begin{cases} 1, & 0 \le t < \frac{1}{2} \\ -1, & \frac{1}{2} \le t \le 1 \\ 0, & \text{otherwise} \end{cases}.$$

The basis function for the approximation was the Haar-scaling function. The corresponding wavelet function is the Haar wavelet given by

$$\psi(t) = \begin{cases} 1, & 0 \le t < \frac{1}{2} \\ -1, & \frac{1}{2} \le t \le 1 \\ 0, & \text{otherwise} \end{cases}.$$

This means that the details are given by

$$g_0(t) = 1 \cdot \psi(t).$$

Note that $g_0(t) \in \mathcal{W}_0$ is orthogonal to $f_0(t) \in \mathcal{V}_0$, like wanted in the construction of $\mathcal{W}_0$.

From $\mathcal{W}_0 \subset \mathcal{V}_1$ and (5.1) follows that the wavelet function $\psi(t)$ can be written as a linear combination of $\phi_{2^j,n}$:

$$\psi(t) = \sqrt{2} \sum_n h_1(n)\phi(2t-n), \quad n \in \mathbb{Z}. \tag{5.5}$$

Equation 5.5 is called ***wavelet equation***. The coefficients h_1 are called ***wavelet function coefficients***. They are a high pass filter and it holds that $\sum_k h_1(k) = 0$ [22]. The wavelet equation (5.5) shows that the wavelet function, which determines the details of f, can be directly calculated from the scaling function. Also the coefficients can be calculated from the scaling function coefficients. In practice the scaling and wavelet function coefficients are commonly chosen finite. In this case it holds that for h_1 with length N:

$$h_1(n) = (-1)^n h_0(N-1-n). \quad [3]$$

According to (5.3) a function $f(t) \in L^2(\mathbb{R})$ can now be rewritten by deriving first its approximation on a coarse scale with the scaling function $\phi(t)$ and then adding the lost details with the wavelet function $\psi(t)$. That means we have

$$f(t) = \underbrace{\sum_k a_0(k)\phi(t-k)}_{\text{approx. at coarse scale } (j=0)} + \underbrace{\sum_j \sqrt{2^j} \sum_k d_j(k)\psi(2^j t - k)}_{\text{details}}. \tag{5.6}$$

For the example this means:

$$f(t) = 2 \cdot \phi(t) + 1 \cdot \psi(t).$$

Now the aim from the beginning of this section is reached: A function is transformed in a way that it is first captured at a coarse scale and then the details are determined by going on finer scales.

In most applications the original function f is a discrete signal either because it is sampled or its discrete from the beginning, like an image. That means the highest resolution is equal to the sample level. The coefficients $a_j(k)$ and $d_j(k)$ that are needed for the wavelet expansion form the wanted ***discrete wavelet transform***. Like mentioned before, a closed from of the scaling function ϕ is not needed to determine the coefficients a_j. The same holds for the wavelet function ψ and d_j. Instead the coefficients can be derived without the use of ϕ and ψ using filterbanks [17]. This will be described in the next section.

5.2 Filterbank Implementation of the Discrete Wavelet Transform

Mallat showed, that it is possible to compute the coefficients a_j and d_j of equation (5.6) by a convolution followed by downsampling. The coefficients a_j which are needed for the calculation of the approximation of the original function at resolution 2^j can be determined by a convolution of a_{j+1} with $\hat{h}_0$ and downsampling the output by two [17]. $\hat{h}_0$ is the mirror filter of the scaling function coefficients h_0, i.e. $\hat{h}_0(n) = h_0(-n)$. The same holds for the detail coefficients, which can be computed by convolving the coefficients a_{j+1} with the mirror filter of the wavelet coefficients $\hat{h}_1$ and downsampling the output by 2. In formulas this means

$$a_j(k) = (\hat{h}_0 * a_{j+1})(k) \downarrow 2,$$
$$d_j(k) = (\hat{h}_1 * a_{j+1})(k) \downarrow 2,$$

where ↓ 2 denotes the downsampling. This equations can be rewritten in terms of the scaling function coefficients h_0 and wavelet function coefficients h_1 by writing out the convolution equation

$$a_j(k) = \sum_m \hat{h}_0(2k-m)a_{j+1}(m) = \sum_m h_0(m-2k)a_{j+1}(m), \tag{5.7}$$

$$d_j(k) = \sum_m \hat{h}_1(2k-m)a_{j+1}(m) = \sum_m h_1(m-2k)a_{j+1}(m). \tag{5.8}$$

The filter h_0 is a low pass filter, so the coefficients a_j give the averages at resolution 2^j, the filter h_1 on the other hand is a high pass filter, so d_j give the differences at resolution 2^j.

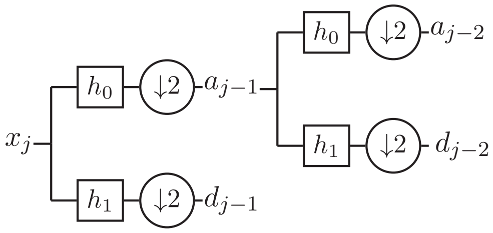


Figure 5.2: Filterbank diagram of two decomposition steps of the 1D discrete wavelet transform.

With equations 5.7 and 5.8 it is possible to compute the discrete wavelet transform of a signal. Let x be the original one dimensional discrete signal of length N with $N = 2^j, j \in \mathbb{N}$. For example $x = [3,5,1,8]$, with $N = 4 = 2^2$. As an example this signal is now decomposed with the Haar wavelet. The corresponding scaling and wavelet coefficients are given with

$$h_0 = \frac{1}{\sqrt{2}}[1,1],$$
$$h_1 = \frac{1}{\sqrt{2}}[1,-1].$$

The filterbank diagram of the discrete wavelet transform till the level 2 is shown in figure (5.2). First, the original signal at resolution 2^j is decomposed into average coefficients a_{j-1} and detail coefficients d_{j-1} at the next coarser resolution 2^{j-1}. In the example this coarser resolution is 2^1. The average coefficients are calculated with $a_2 * \hat{h}_0 = a_2 * h_0 = \frac{1}{\sqrt{2}}[3,8,6,9,8]$, were $a_2 = x$. From this every second value is taken in the downsampling step, which means that $a_1 = \frac{1}{\sqrt{2}}[8,9]$. The detail coefficients at this resolution, d_1, are calculated analog with the filter h_1. The whole decomposed signal at this resolution is $[a_1, d_1] = \frac{1}{\sqrt{2}}[8,9,-2,7]$.
With every convolution step, the length of the signal is halved. In the second step the coefficients a_{j-1} are decomposed further into average and detail coefficients at the next coarser scale. In the example this second decomposition is given with $[a_0, d_0, d_1] = [\frac{17}{2}, -\frac{1}{2}, -\frac{2}{\sqrt{2}}, \frac{7}{\sqrt{2}}]$. The signal can be decomposed until $j = 0$ where just one coefficient a_0 remains, like in the second decomposition in the example.
With the detail coefficients the whole original signal x can be reconstructed from this a_0, like in equation (5.6). This can also be done with a filterbank implementation. The reconstruction is not discussed here, because it is not needed for the development of a feature. But for example in image compression, the most popular application of the dis-

crete wavelet transform, reconstruction and in particular the term *perfect reconstruction* play an important role.

5.2.1 Boundary Conditions

By convolving a signal with a finite filter, the boundary values of the signal $x(0), ...,$ $x(N-1)$ must be considered. For example the last average coefficient of the first decomposition $a_{j-1}(\frac{N}{2}-1)$ is due to equation (5.7) calculated with

$$\begin{aligned} a_{j-1}(N/2-1) &= \sum_m h_0(m-(N-2))a_j(m) \\ &= h_0(0)a_j(N-2) + h_0(1)a_j(N-1) + h_0(2)a_j(N) + h_0(3)a_j(N+1), \end{aligned}$$

where $a_j = x$. But $a_j(N)$ and $a_j(N+1)$ are not defined because x is a finite signal. To calculate $a_{j-1}(\frac{N}{2}-1)$ a boundary condition has to be defined. In this work symmetric extension is used as boundary condition. That means the last values of x a mirrored. Thus, the value $a_{j-1}(\frac{N}{2}-1)$ can be calculated with

$$\begin{aligned} a_{j-1}(\frac{N}{2}-1) = {} & h_0(0)a_j(N-2) + h_0(1)a_j(N-1) + h_0(2)a_j(N-1) \\ & + h_0(3)a_j(N-2). \end{aligned}$$

Analogous the last detail coefficient $d_j(\frac{N}{2}-1)$ is calculated with

$$\begin{aligned} d_{j-1}(\frac{N}{2}-1) = {} & h_1(0)a_j(N-2) + h_1(1)a_j(N-1) + h_1(2)a_j(N-1) \\ & + h_1(3)a_j(N-2). \end{aligned}$$

This condition is used because it is easy so implement. Furthermore it does not produce artifacts at the end of the transformed signal, which can occur for example with zero padding.

5.2.2 Two Dimensional Transform

With the described discrete wavelet transform it is possible to transform one dimensional signals. But in most applications and also in this work, the aim is to transform images, which are two dimensional. If the scaling function, and with is the whole multiresolution approximation, is separable, the 1D theory of the discrete wavelet transform can easily be transferred into the 2D situation. Because this it is very similar to

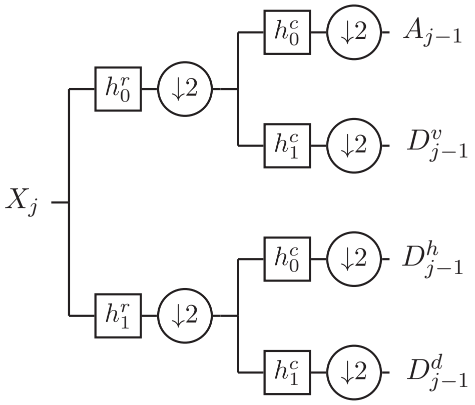
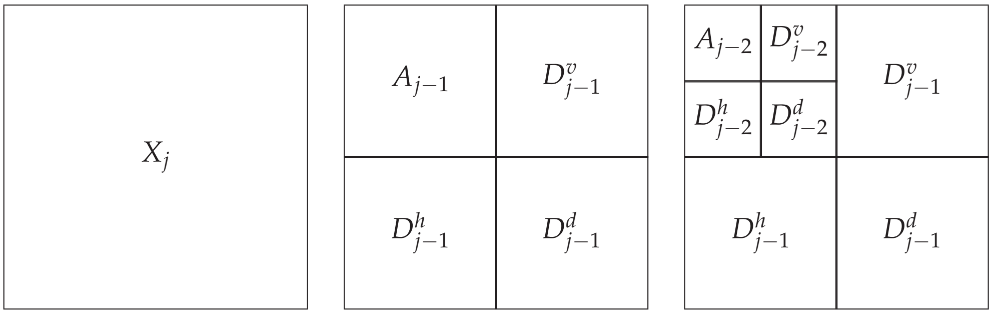


Figure 5.3: Filterbank diagram of the first decomposition level of the 2D discrete wavelet transform.

the 1D case, the theory for the two dimensional discrete wavelet transform is skipped here. It can be found for example in [17]. Here only the two dimensional filterbank implementation is discussed.

The filterbank design of the discrete wavelet transform of images is shown in figure 5.3. It consists of one dimensional transforms, performed first on the rows and afterwards on the columns of the image. The input image is denoted as X with size $N \times N, N = 2^j$. First the rows of X are low-pass filtered with the one dimensional filter h_0 which is denoted by h_0^r in the filterbank diagram and high-pass filtered with h_1, denoted by h_1^r. Afterwards the same is done with the columns, but for each part, the low-pass filtered as well as the high-pass filtered, separately. It follows that the transformed image contains four different filtered versions of the original image. One part, denoted by A, is low-pass filtered on rows and columns. That means it contains the averages of the original image on a coarser resolution. Three parts, denoted by D^v, D^h and D^d, give the details of the original image. The size of the transformed image stays the same like the size of the original image. This is due to the fact that after every convolution and downsampling the length of a vector halves, as described in the one dimensional case. In the 2D case the convolution plus downsampling is performed on the rows *and* the columns. That means that the parts of the first transformation level $A_{j+1}, D_{j+1}^v, D_{j+1}^h$ and D_{j+1}^d have the size $\frac{N}{2} \times \frac{N}{2}$. The whole transformed image, which consists of this four parts has again size $N \times N$. This is illustrated in figure 5.4.



orignial 1. transformation 2. transformation

Figure 5.4: Schema of the 2D discrete wavelet decomposition

Lets now come to the interpretation of the two dimensional discrete wavelet transform. The four parts of the decomposition A, D^v, D^h and D^d all contain different information of the original image. The subband A contains the averages, like already discussed. This is because a low-pass filter lets low frequencies pass and attenuates higher frequencies. On the other hand a high-pass filter lets high frequencies pass and attenuates the low ones. If the high-pass is only performed on the columns, the vertical high frequencies are passed by the filter. This means that the subband D^v gives the vertical high frequencies and horizontal low frequencies, because the rows are low-pass filtered and the columns high-pass filtered. Thus the D^v part of the transformed images highlights the vertical edges of the original image. Analogous the D^h part gives the horizontal edges and D^d the high frequencies in both vertical and horizontal directions, which correspond to diagonal edges. This is shown with a simple test image in figure 5.5.
This interpretation of the two dimensional discrete wavelet transform is important for the understanding of the feature discussed in section 5.4.

5.3 Wavelet Bases

In this section the choice of the scaling function $\phi(t)$ and the wavelet function $\psi(t)$ is discussed. Closed forms of these functions are not needed for the calculation of the discrete wavelet transform, as shown in the previous sections. Furthermore, characteristics of the scaling function as well as the wavelet function can be determined indirectly

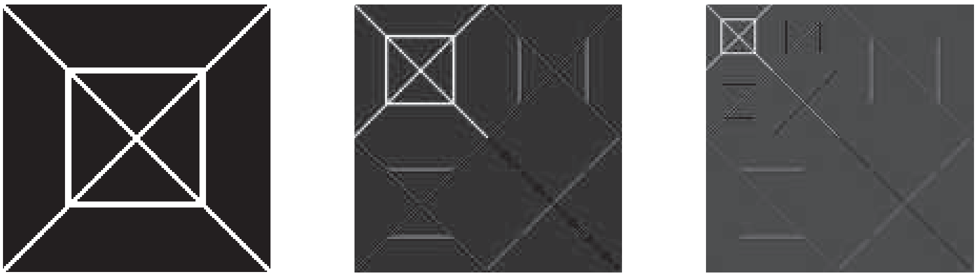


Figure 5.5: 2D discrete wavelet transform of a test image. On the left hand side the original image is shown, in the middle the first level decomposition and on the right the second level. The image is transformed with Daubechies D4 filter, see section5.3.1.

through properties of the corresponding coefficients h_0 and h_1.

In the theory of the wavelet transform described above, the wavelet and scaling function have to be orthogonal in order that a signal can be decomposed without loss of information. A family of wavelets constructed by I. Daubechies with this property is presented in the first part of this section. However, in the application of wavelets in feature extraction symmetric wavelets are often used. Livens pointed out in [14], that symmetry is an important property for a wavelet basis, because with it is at least guaranteed that the same feature is obtained if the image is turned upside down. With biorthogonal wavelets it is possible to achieve this property. They are described in section 5.3.2.

In the last part of this section the wavelet bases used in this work to built a texture feature are given.

5.3.1 Orthogonal Wavelets

Ingrid Daubechies derived scaling function coefficients that produce orthonormal wavelets with compact support and a high number of vanishing moments [6]. The number of vanishing moments of a wavelet function ψ indicates which polynomials can be completely represented by the scaling function, without any loss of information. For example, this is important in the application of wavelets in image compression. For the construction of wavelet based features it is more important that the wavelets have compact support. From this property follows that the scaling and wavelet function coefficients h_0 and h_1 are finite. This is necessary for the implementation of the discrete wavelet

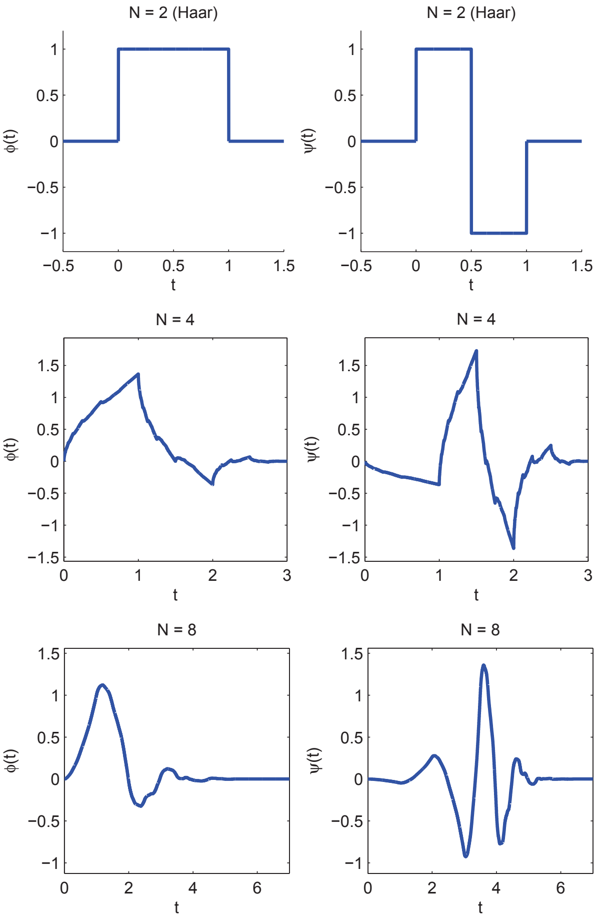


Figure 5.6: Daubechies Scaling- and Wavelet function for different N.

transform.
Daubechies constructed different scaling functions with N coefficients h_0 and $N/2$ vanishing moments. Some examples are shown in figure 5.6. The wavelet with $N = 2$ is also known under the name Haar-wavelet. The corresponding coefficients for $N = 2$ and $N = 4$ are

- $N = 2$ (Haar-Wavelet):
$$h_0 = \tfrac{1}{\sqrt{2}}(1, 1)$$
$$h_1 = \tfrac{1}{\sqrt{2}}(1, -1)$$
- $N = 4$:
$$h_0 = \tfrac{1}{4\sqrt{2}}(1 + \sqrt{3}, 3 + \sqrt{3}, 3 - \sqrt{3}, 1 - \sqrt{3})$$
$$h_1 = \tfrac{1}{4\sqrt{2}}(1 - \sqrt{3}, -3 + \sqrt{3}, 3 + \sqrt{3}, -1 - \sqrt{3})$$

These filters are examples, which can be used for an orthogonal multiresolution analysis. In the orthogonal case the same filters are used for decomposition and reconstruction, thus they cannot be symmetric, except for $N = 2$ [5]. If symmetric filters are wanted, one has to give up orthogonality. This is described in the next section.

5.3.2 Biorthogonal Wavelets

With biorthogonal bases is it possible to construct symmetric wavelets. They were developed by Cohen, Daubechies and Feaveau [5] and are commonly called CDF-wavelets.
Symmetry can not be achieved by using the same filters for decomposition and reconstruction like in the orthogonal case. Therefore dual pairs of scaling and wavelet functions are constructed. That means different filters are used for decomposition and reconstruction. Although the reconstruction is not needed in the particular application of wavelets as feature, it is also mentioned at this point, because it is important in the development of symmetric wavelets.

The multiresolution theory of the orthogonal case has to be modified to get symmetric wavelets. The multiresolution vector spaces needed for reconstruction of the signal now differ from the ones needed for decomposition. The approximation of the signal is done in the vector spaces $\{\tilde{\mathcal{V}}_{2^j}\}_{j\in\mathbb{Z}}$, the reconstruction in the vector spaces $\{\mathcal{V}_{2^j}\}_{j\in\mathbb{Z}}$. The bases of this vector spaces are given by $\tilde{\phi}_{2^j,k}$ and $\phi_{2^j,k}$. The same holds for the calculation of the details in $\{\tilde{\mathcal{W}}_{2^j}\}_{j\in\mathbb{Z}}$ and $\{\mathcal{W}_{2^j}\}_{j\in\mathbb{Z}}$ respectively. They have the bases $\tilde{\psi}_{2^j,k}$ and $\psi_{2^j,k}$. Again, $\tilde{\mathcal{W}}_{2^j}$ is a complement of $\tilde{\mathcal{V}}_{2^j}$ in $\tilde{\mathcal{V}}_{2^{j+1}}$ and $\mathcal{W}_{2^j}$ is a complement of $\mathcal{V}_{2^j}$ in $\mathcal{V}_{2^{j+1}}$, but this time they are not orthogonal [5]. Instead each scaling space is orthogonal to the dual wavelet space:

$$\mathcal{V}_{2^j} \perp \tilde{\mathcal{W}}_{2^j} \quad and \quad \mathcal{W}_{2^j} \perp \tilde{\mathcal{V}}_{2^j}. \quad [22]$$

Again, the scaling and wavelet functions are constructed using the filter coefficients. The coefficients are related to the functions with:

$$\begin{aligned}\tilde{\phi}(t) &= \sqrt{2}\sum_n h_0\tilde{\phi}(2t-n), \quad \tilde{\psi}(t) = \sqrt{2}\sum_n h_1\tilde{\psi}(2t-n),\\ \phi(t) &= \sqrt{2}\sum_n f_0\phi(2t-n), \quad \psi(t) = \sqrt{2}\sum_n f_1\psi(2t-n). \quad [5]\end{aligned}$$

The filters h_0, h_1 for decomposition and f_0 and f_1 for reconstruction are chosen in a way that perfect reconstruction is possible. Therefore it must hold that

$$\begin{aligned}h_1(n) &= (-1)^{n+1}f_0(-n),\\ f_1(n) &= (-1)^{n+1}h_0(-n). \quad [5]\end{aligned}$$

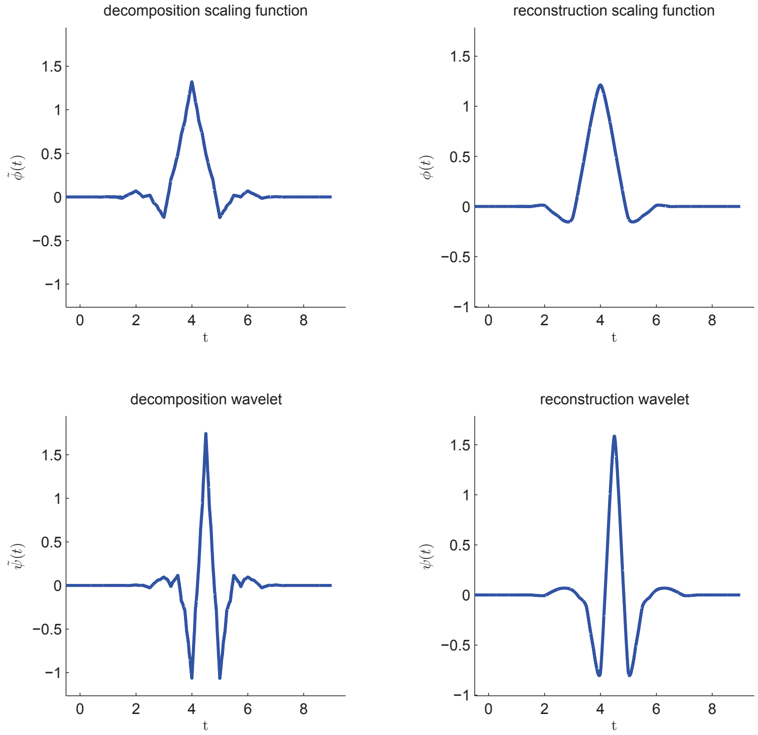
The properties of the dual scaling and wavelet functions are again determined with the filter coefficients. The dual CDF-wavelets $\tilde{\psi}$ and ψ are built in a way that they have compact support and $\tilde{p}$ and p vanishing moments. Additionally these wavelets can be symmetric. The filters with four vanishing moments in decomposition and reconstruction are given in table 5.1 [5]. In this work they are denoted with 'biorthogonal 4.4'. The resulting scaling and wavelet functions are given in figure 5.7.

5.3.3 Choice of the Wavelets Basis for Feature Extraction

In this work three different wavelet bases were tested regarding to their suitability as basis for the wavelet features presented in the next section. The first one is the Haar wavelet because it just contains two filter coefficients and is therefore of low complexity. As second basis the Daubechies wavelet with 4 vanishing moments (D4) is tested. It has better approximation characteristics than Haar, whereas the filter length is still

n	$h_0/\sqrt{2}$	$h_1/\sqrt{2}$
0	0.602949018236	-0.557543526229
1,-1	0.266864118443	0.295635881557
2,-2	-0.078223266529	0.028771763114
3,-3	-0.016864118443	-0.045635881557
4,-4	0.026748757411	0

Table 5.1: Decomposition filter coefficients of biorthogonal wavelets with four vanishing moments $\tilde{\psi}$ and ψ. They can be also found in [5].


decomposition scaling function

reconstruction scaling function

decomposition wavelet

reconstruction wavelet

Figure 5.7: Dual scaling and wavelet functions of biorthogonal 4.4 wavelets.

low. As the third and final tested basis a symmetric wavelet is chosen with 4,4 vanishing moments, denoted as biorthogonal 4.4. This wavelet basis is often used to build a texture feature, for example in [12]. The results of the comparison of the different bases can be found in chapter 6.

5.4 Wavelet Based Features

After the theory and implementation of the discrete wavelet transform were explained in the previous sections, now the wavelet transform as feature will be discussed. Theoretically one can use the whole wavelet coefficients as feature vector, but this would imply a huge feature vector and a decreasing classification accuracy. Instead the features should be build in a way that they highlight characteristic properties of the different texture classes whilst simultaneously the dimensionality is kept low.

A common feature is the l_1-norm of the detail coefficients. It is for example used in [12] and [13]. In this work two features a build using the l_1-norm of the detail coefficients, which were also used in a similar way in [12].
First, the image is decomposed J times up to the resolution 2^{j-J}, where $N = 2^j$ is the sampling resolution. The decomposition results in 6 different detail coefficient images for each orientation denoted by D_i^h, D_i^v and $D_i^d (i = 1, ..., 6)$, here the i indicates the decomposition level. For each of these detail images, the l_1-norm, i.e. the mean absolute coefficient (MAC), is calculated with

$$MAC(o,i) = l_1(D_i^o) = \sum_{k_x,k_y} \left| d_i^o(k_x,k_y) \right|,$$

with $\quad i = 1, ..., 6$, $o \in \{h, v, d\}$ and where $d_i^o(k_x, k_y)$ denotes the image value of D_i^o at position (k_x, k_y).
With this 2 different feature sets are build:

$$f_1(i) = \sum_o MAC(o,i),$$

$$f_2(i) = |MAC(h,i) - MAC(v,i)| + MAC(d,i) + \frac{MAC(h,i) + MAC(v,i)}{2},$$

with $i = 1, ..., 6$ and $o \in \{h, v, d\}$.

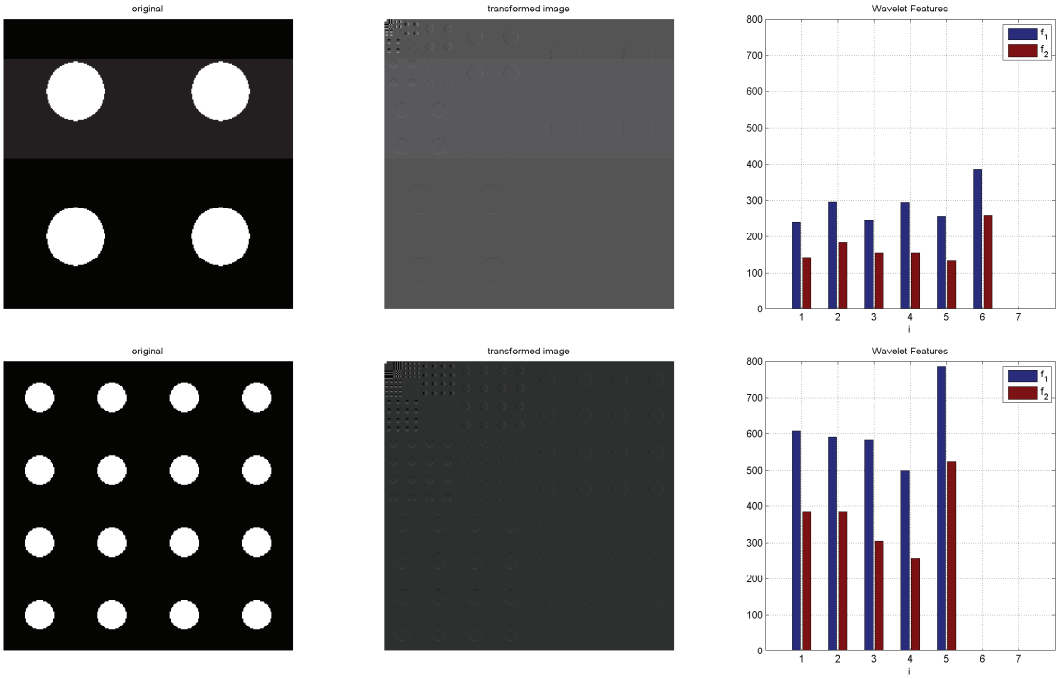


Figure 5.8: Discrete Wavelet Transform of two test images containing circles with different sizes and the resulting wavelet features f_1 and f_2.

The first feature f_1 captures the mean absolute coefficient of each decomposition level i, i.e. the sum of the detail information lost by going down on this level. This lost information contains basically edges, which can also be explained with the calculation of the detail coefficients with low- and high-pass filters, see 5.2.2. With increasing decomposition level, the amount of lost information changes depending on the appearance of the image on the particular resolution. This can be seen in figure 5.8. Two test images containing circles with different sizes are shown in the first column. Their discrete wavelet transform calculated with the Haar basis and corresponding wavelet features are shown in the next columns. The wavelet features show two main differences. First, the values are higher for the second image. This can be explained with the higher amount of edges in this image. Furthermore, the features for both images are zero up to a certain decomposition level. This is because at that resolution the images are sampled at such a coarse scale that the resulting images are completely white and contain no edges. For the first image the features are zero after the sixth decomposition. For the second images the whole information is already represented with five decompositions. Afterwards there are no details left and the features are zero.

The second feature f_2 describes anisotropic structures. This is due to the fact, that each of the detail coefficients and with it their MAC varies in a different way if the main orientation of the image structure changes. This is illustrated with four test images in figure 5.9. The test images are in the first column, their decomposition in the second and the resulting MACs and features f_1 and f_2 in the last two columns. The first test image at the top is not anisotropic, it contains circles. The other three contain lines, which are oriented in horizontal, vertical or diagonal direction. One can see in the decomposition images that on each level the detail coefficients correspond to a particular direction. This is due to the two the different filters, as already explained in the previous section. It follows that also the MACs of the detail coefficient image parts differ. Horizontal lines lead to high $MAC(h,i)$, while $MAC(v,i)$ and $MAC(d,i)$ are low. Vertical lines on the other hand lead to low $MAC(h,i)$ and $MAC(d,i)$ and high $MAC(v,i)$. This affects the feature f_2, in both cases the first summand increases. With diagonal lines the MACs of all directions increase, whereas the values of $MAC(h,i)$ and $MAC(v,i)$ are very similar. Thus, the first summand gets very small and the last summands are increasing. All in all the feature f_2 should increase if the structure in the image has a preferred direction.

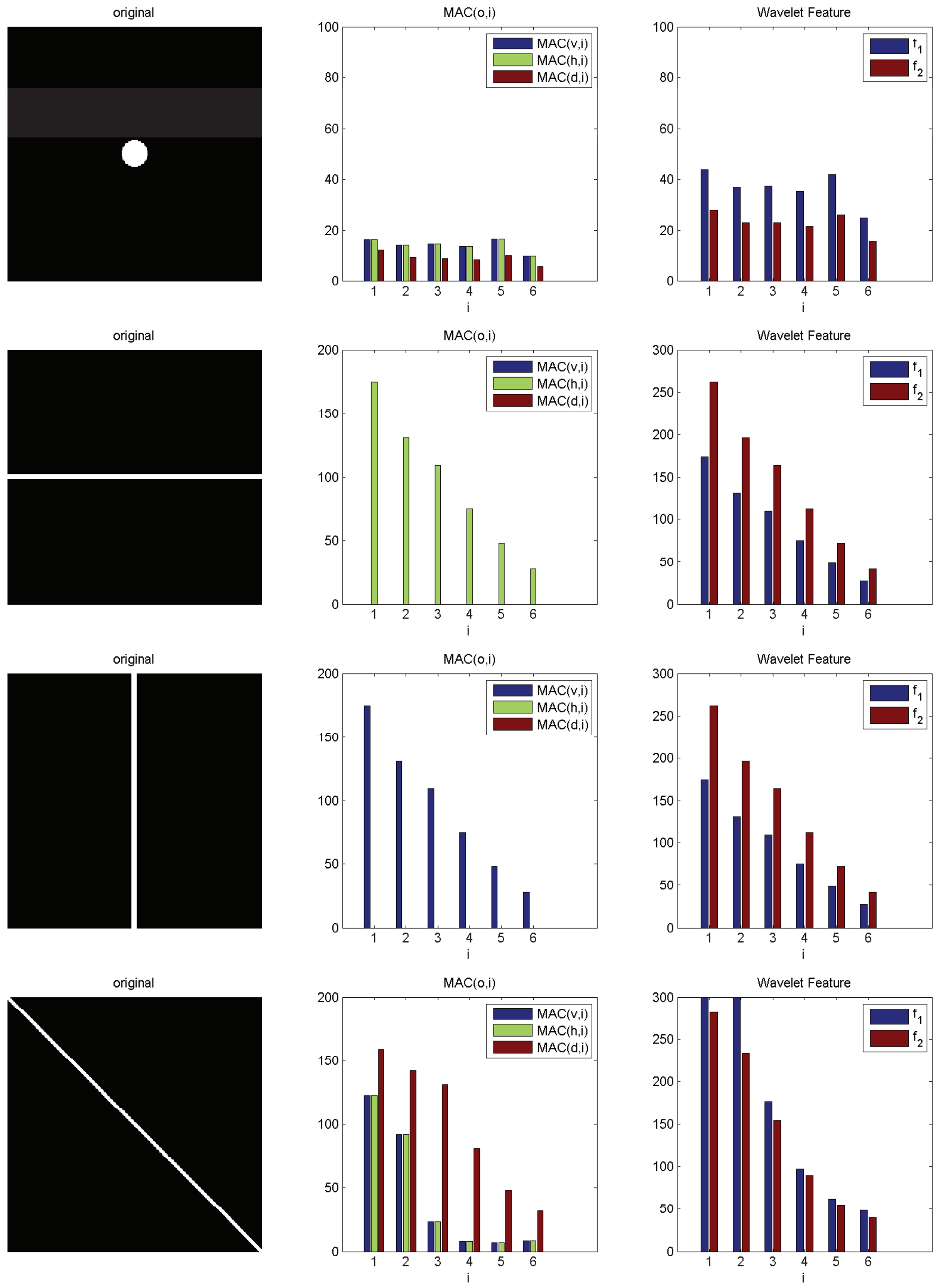


Figure 5.9: MACs and the resulting wavelet features f_1 and f_2 for different test images.

Chapter 6

6 Results

6.1 Feature Evaluation

In this section the features presented in the previous sections are compared and analyzed. First the choice of the wavelet basis is evaluated by comparing the wavelet features calculated with Haar, Daubechies 4 or biorthogonal 4.4 wavelet basis. Secondly the intensity attributes, local binary patterns and wavelet features for two example images of adenocarcinoma subtypes are shown and compared. The wavelet features are additionally interpreted for other classes. Afterwards the discriminative power of the three feature types are compared by classifying the microscopic lung tumor tiles with just one feature type at once.

6.1.1 Evaluation of the Basis for the Wavelet Feature

Like mentioned in 5.3.3, three wavelet bases were tested: Haar wavelets, Daubechies wavelets with four vanishing moments and biorthogonal wavelets with four vanishing moments for the decomposition and reconstruction wavelet. The wavelet features were calculated with each of the three bases for 20 image regions of every tumor class. Additionally the features were calculated for the alveolar lung tissue class (in the plots named ’normal’) and fibrous tissue because these two classes are the most common healthy tissue classes that occur next to the tumor classes. The size of the test images was 1024 × 1024 pixel on the highest magnification. This size is equal to the tile size of

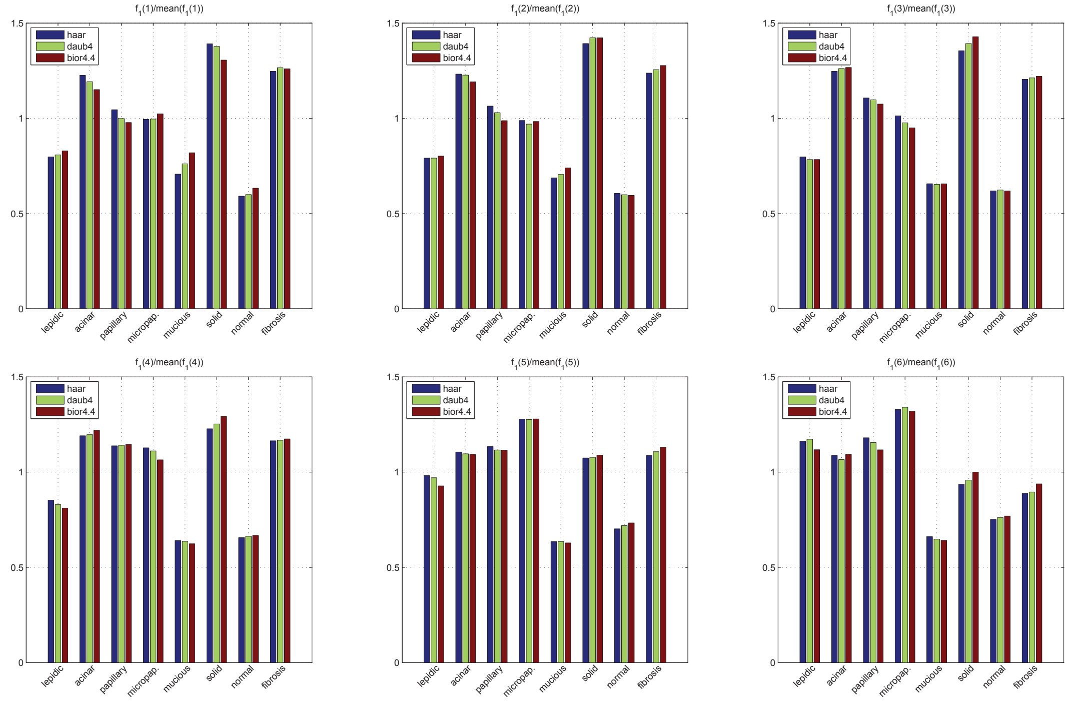
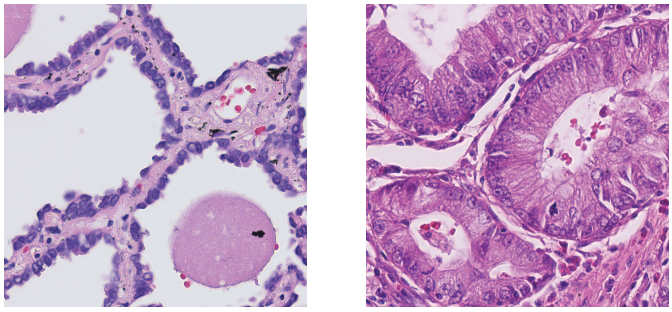


Figure 6.1: Comparison of the wavelet feature f_1 with Haar, Daubechies 4 and biorthogonal 4.4 wavelet basis. The normalized mean value of the feature for 20 test images for each class is shown.



(a) Lepidic adenocarcinoma **(b)** Acinar adenocarcinoma

Figure 6.2: Example images of lepidic and acinar adenocarcinoma.

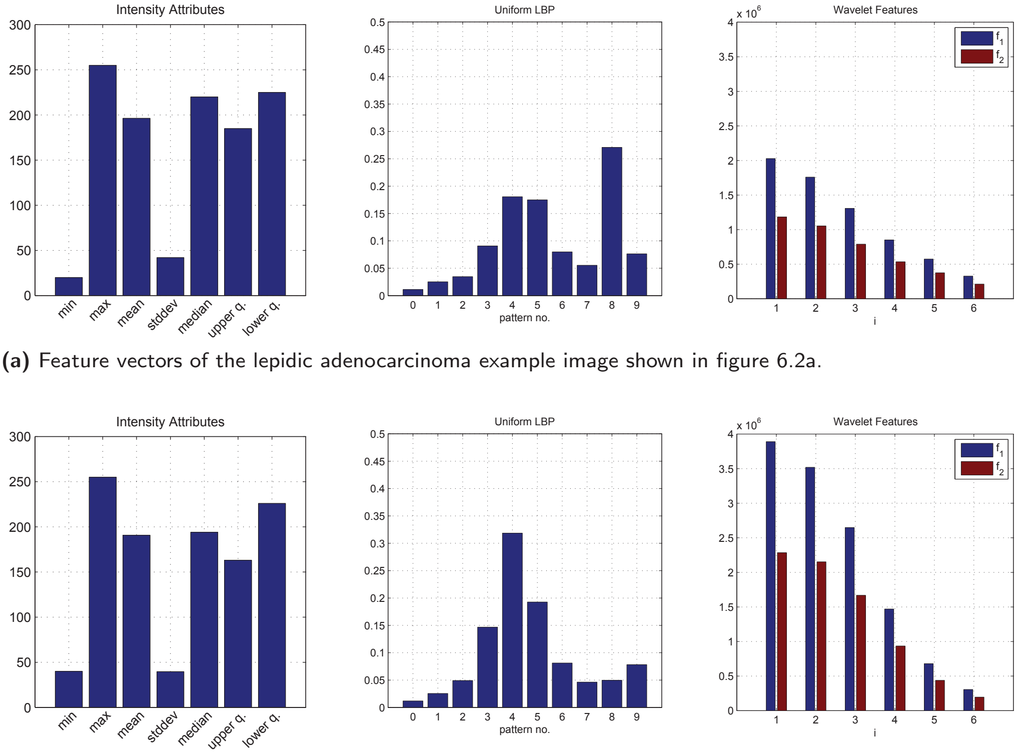
the final classification with HistoCAD.

The results of f_1 till the decomposition level $i = 3$ are shown in figure 6.1. In this figure the mean of the 20 test images for each class is plotted. Additionally these values are normalized in order to be comparable. The plots show, that the differences between the wavelet bases are not significant. For each basis the resulting feature is similar for a particular class. The same holds for the second feature f_2 and if the images are decomposed further. Because of these results the Haar basis is used from here on. It has the smallest filter length and has therefore the lowest complexity.

6.1.2 Comparison of Different Features on Histological Images of Lung Tumors

As an example, the resulting feature vectors of the intensity, LBP and wavelet attributes of tiles of lepidic and acinar lung adenocarcinoma are shown and discussed now. The image tiles are shown in figure 6.2. As described in chapter 2 lepidic adonocarcinoma still appears differentiated. Tumor cells are spread on the walls of the alveoli without destroying them. The acinar subtype on the other hand consists of glandular structures with a central lumina. The differences of this two subtypes are clear visible by the human eye.

The resulting feature vectors of the example images calculated on the red color chan-



(a) Feature vectors of the lepidic adenocarcinoma example image shown in figure 6.2a.

(b) Feature vectors of the acinar adenocarcinoma example image shown in figure 6.2b.

Figure 6.3: Comparison of intensity attributes, local binary patterns and wavelet based features for lepidic and acinar adenocarcinoma.

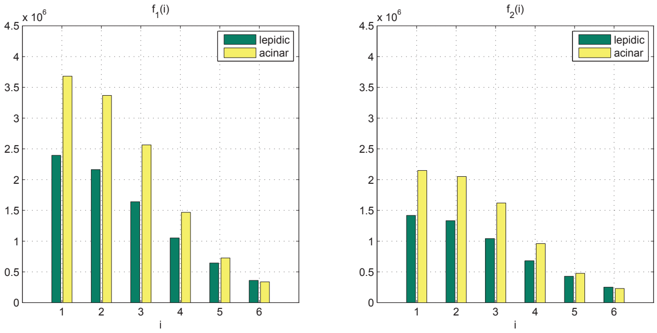


Figure 6.4: Direct comparison of the wavelet features for lepidic and acinar adenocarcinoma.

nel are shown in figure 6.3. The wavelet based features f_1 and f_2 are in this example calculated with the Haar basis and the images are decomposed 6 times.
Differences between the feature vectors of the two subtypes can be seen in all three cases. Because the lumina of the alveoli are not filled with tumor cells in the lepidic subtype. This image has more bright parts than the image with acinar adenocarcinoma. This is also reflected in the intensity attributes. The values for median, lower and upper quartile are higher in the image with lepidic tissue. Furthermore the pattern #8 of the local binary patterns occurs more often for this image. This pattern detects bright spots or areas. On the other hand the occurrences of pattern #3, #4 and #5 are significantly increased for the acinar subtype. These three patterns recognize edges in an image, where the intensity of the edge is neglected. The high occurrence of these patterns can be explained with higher density of the tissue which goes along with more color changes.

By looking at the wavelet features of the two images one recognizes the increasing decomposition level goes along with decreasing feature values. This is the result of the coarsening of the resolution. With a coarser resolution the details of the images increase which results in decreasing values f_1 and f_2. The differences between the wavelet feature of lepidic and acinar tissue can be seen more clearly in figure 6.4. In these plots the mean values of f_1 and f_2 for 20 example images of the two tumor classes are shown. Generally the wavelet features for the acinar subtype are higher than the ones for the lepidic adenocarcinoma. This is again because the lepidic tissue contains less structure

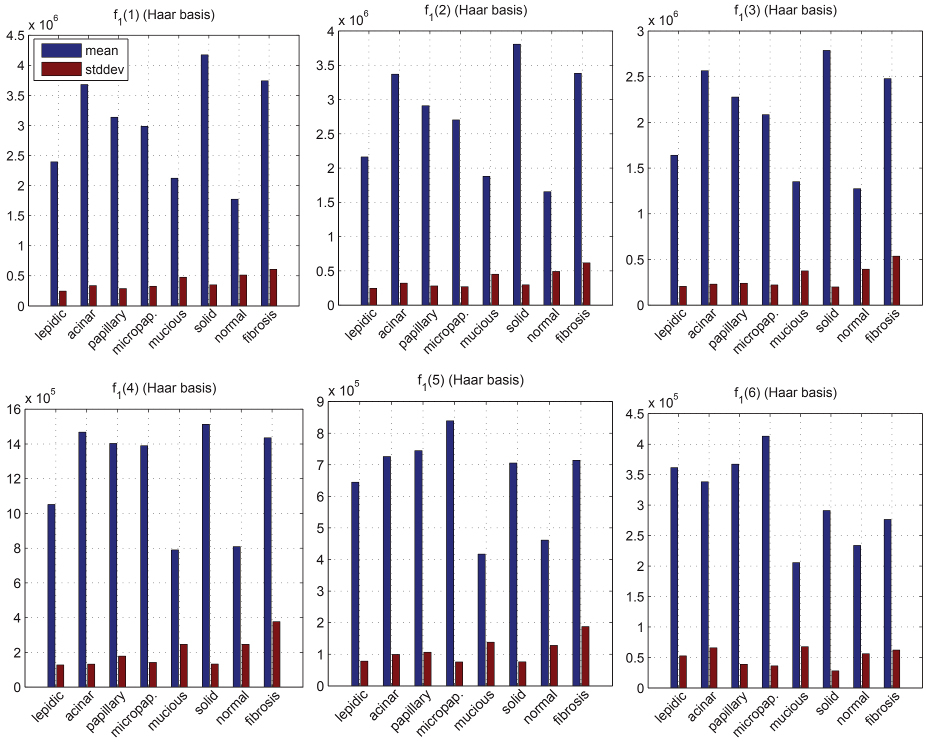


Figure 6.5: Predictive power of wavelet based features. Mean and standard deviation for f_1 for eight different classes.

because of the remaining alveolar lumina. With increasing decomposition level the values decrease in both cases, whereas the values for the lepidic subtype decrease slower. For the decomposition level $i = 6$ the feature in the lepidic case are actually higher than in the acinar case. This is due to the different kinds of edges in the two images. The still recognizable alveolar walls in the lepidic subtype are wide edges, that are also captured at a coarser resolution. The edges in the acinar subtypes on the other hand are finer, because they originate mainly from slight changes in the color. They are not captured at a coarse scale. The observations in the development of the wavelets features hold for the feature f_1 as well as for f_2. To determine the differences between these feature now the results for more classes are compared.

Altogether the wavelet features f_1 and f_2 were calculated till the decomposition level $i = 6$ for 20 example images of each of the six tumor classes and additionally for the classes fibrosis and normal alveolar lung parenchyma. The resulting mean and stan-

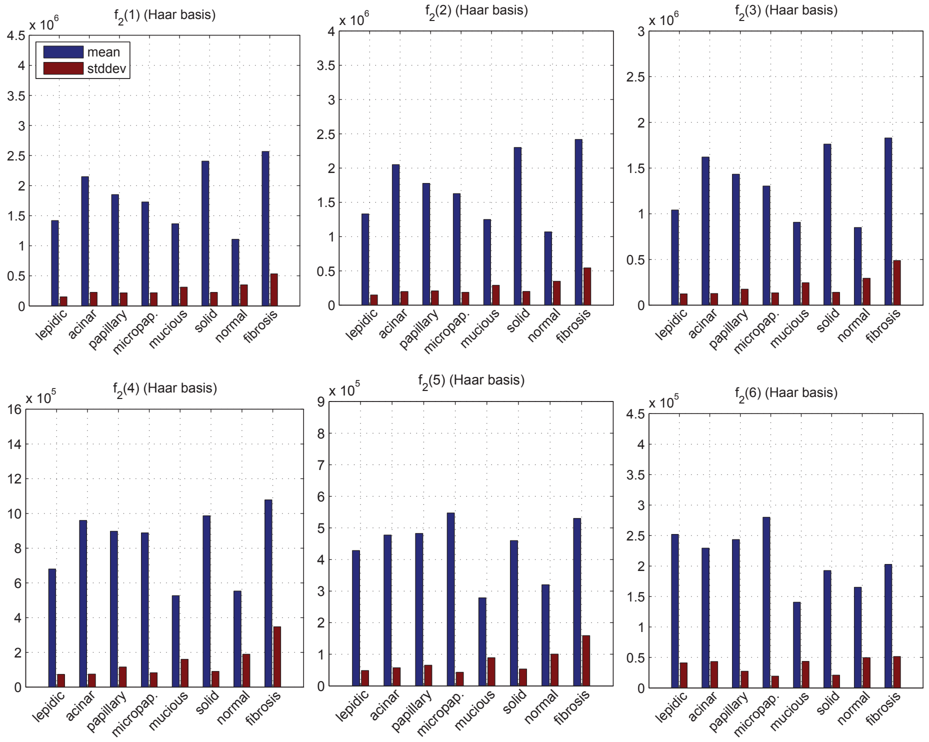


Figure 6.6: Predictive power of wavelet based features. Mean and standard deviation for f_2 for eight different classes.

dard deviation are shown in figure 6.5 (f_1) and 6.6 (f_2). f_1 reaches on the first decomposition level the highest values for the classes acinar and solid adenocarcinoma, and fibrosis. These three classes all contain tissue with a high cell density, which explains the high feature values. On the coarsest resolution on the other hand the highest values for f_1 belong to lepidic, papillary and micropapillary adenocarcinoma. These are tissues which tend to have more luminal spaces between the cell structures and have therefore edges which are more outstanding and can also be recognized on a coarser resolution. This is also the case for normal alveolar lung parenchyma. Although the values for this tissue class are comparatively low for each i, they increase for example relatively to the values of solid adenocarinoma. All in all it is possible to discriminate these classes with the wavelet feature f_1.

The comparison of f_2 for the eight classes is shown in figure 6.6. This feature was build to detect structures that are oriented in a preferred direction, like for example fibrous tissue. In the results can be seen, that for $f_2(1)$-$f_2(4)$ the highest values belong to fibrous tissue. Relatively to the other classes the values for this class are higher, but in consideration of the variances the differences to f_1 are low. That means the property that f_2 detects anisotropic structures cannot be confirmed in this test.

6.1.3 Classification Accuracy with one Feature Type

In this part it is tested how much classification accuracy can be achieved if just one of the three image features is used. That means the accuracy of a classification just with intensity attributes, local binary patterns or wavelet features is determined and the results are compared. This is done to determine which feature type is important for a good classification. The tests are solely based on an evaluation of example data with Rapid Miner.

The example data set contains 11.976 tiles that are selected uniformly distributed from the ground truth of the 22 histological images of lung adenocarcinoma. The intensity, local binary pattern or wavelet feature vectors are calculated on each of the RGB color channels on different magnifications. Additionally the feature vectors are calculated on the three HSV color channels, because this color space is not as correlated as the RGB space. Additionally, hue, saturation and intensity may give more relevant information of the images. All in all the features are calculated on 6 color channels for 6×, 12×, 25× and 50× magnification which correspond to images sizes of 16×16, 32×32, 64×64 and 128×128 pixels. Higher magnifications are omitted, because the calculation time would be too high. The wavelet features are calculated till the decomposition level $i = 3$ for 6×, $i = 4$ for 12×, $i = 5$ for 25× and $i = 6$ for 50× magnification.

With this data set the performance of each attribute is determined with Rapid Miner as follows. First a feature selection procedure is performed on the example data set using correlation based feature selection in combination with forward selection for intensity attributes, local binary patterns and wavelet features separately. Afterwards, the data set is split into two sets of equal size. One is used for training, the other for testing. The learning algorithm is the random forest implemented in Rapid Miner with

attribute	accuracy	# calculated features	# selected features
intensity attributes	74.0 %	144	29
local binary patterns	71.4 %	210	47
wavelet features	69.7 %	192	48

Table 6.1: Comparison of the classification accuracy of intensity, LBP and wavelets attributes

25 decision trees, and $int(log_2 M + 1)$ features considered at each node to find the best split, where M is the number of features. With the trained algorithm the test set is classified and the accuracy is determined.

The results can be found in table 6.1. It can be seen that the classification just with intensity attributes achieves the best classification accuracy while simultaneously using the smallest feature set. The classification just with features based on local binary patterns or wavelets require larger feature sets but the accuracy is lower. The classification just with local binary patterns produces a little better results. This verifies the assumption of chapter 4, that a lot of information in histological images is given by the color intensity.

6.2 Classification with HistoCAD

In this section the classification of lung tumor tissue with HistoCAD with two feature sets is compared. The first set contains just the attributes that where already available for other classification task in HistoCAD: the intensity attributes and local binary patterns. In the following this feature set is called 'feature set 1'. The second set additionally contains wavelet features, it is named 'feature set 2'. It is compared if a better classification result can be achieved if wavelet features are included.

The feature sets contain all feature vectors calculated on 6 color channels (RGB and HSV) and $6\times$, $12\times$, $25\times$ and $50\times$ magnification, feature set 1 without wavelet features, feature set 2 with. The wavelet features are calculated till the same decomposition levels like in section 6.1.3. On both sets the same feature selection procedure is performed. First, the best subset is found with Rapid Miner using correlation based feature selec-

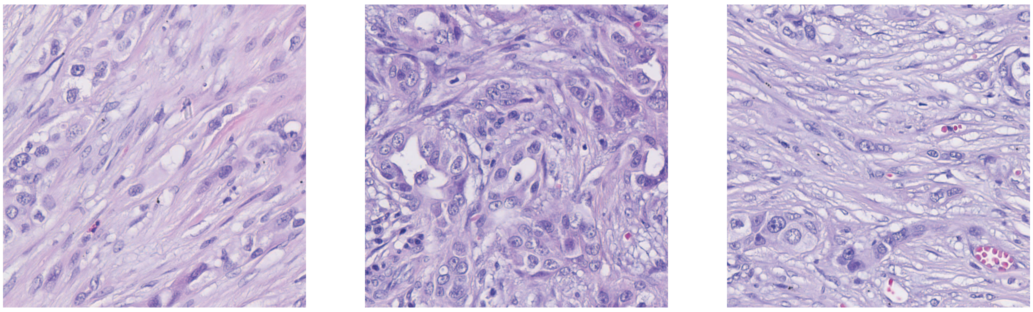
tion in combination with forward selection for an example data set containing 11.976 tiles, like in 6.1.3. During this procedure not all features of a feature vector are selected. Let's take an example: The feature vector of the intensity attributes contains 8 values (min, max, sum, mean, standard deviation, median, lower quartile, upper quartile). But it is possible that just the minimum is selected during the feature selection procedure. Nonetheless the full feature vector has to be calculated. Because of that the selected subset is further reduced, in order that not so many feature vectors have to be calculated and the calculation time is decreased. The classification accuracy decreases in this step, but only as many features are eliminated that the accuracy determined with Rapid Miner decreases not more than 2%.
In the following the best feature subset of feature set 1 is called 'subset 1'. The best subset of feature set 2, which contains wavelet features, is called 'subset 2'.

Each of the resulting subsets is integrated in HistoCAD and the lung tumor data set containing 22 tissue section is classified. The training samples are taken out of the ground truth set by Dr. Frederick Klauschen, where tiles that cannot clearly be assigned to one class are taken out. For some images the ground truth is only available for parts of the image and a full ground truth is only available for 9 of the 22 images. The training example tiles are taken uniformly distributed from this ground truth, were two of the images with full ground truth are left out for evaluation purposes. Altogether the training set contains round about 12.000 tiles.

For some of the images the results of the classification are evaluated with the ground truth that was not used to create the testing set. In the next section some problems are described that have to be taken into account during the validation. The results of the classification are presented and compared subsequently.

6.2.1 Validation Difficulties

The ground truth for the determination of the classification accuracy has to be set by hand. For each tile the class has to be determined. Because this procedure needs a lot of time, a full ground truth is only available for 9 of the 22 images.
Although the ground truth is set by a pathologist, it may contain errors. This has two



(a) Ground truth: fibrous adenocarcinoma

(b) Ground truth: acinar adenocarcinoma

(c) Ground truth: solid adenocarcinoma

Figure 6.7: Three tiles with tissue areas that look very much alike, but are in the ground truth referred to different classes.

main reasons. In some cases a tissue area cannot clearly be assigned to a particular class and the classification depends a lot on the subjective interpretation of the tissue. See for example figure 6.7. The three images look very alike because all of them contain fibrous tissue. But the images 6.7b and 6.7c are assigned to the acinar and solid subtype respectively. Image 6.7b certainly contains lumina that indicate acinar adenocarcinoma as well as image 6.7c contains cells that refer to the solid subtype, but this differences are small. The threshold at which point these images are assigned to one class or another is a subjective decision. At the same time it is not totally false, if all of these three images are classified as fibrous tissue by the classification algorithm, although the ground truth is different.

The second difficulty with the ground truth is that each tile has to be assigned to a particular class, although it may contain different classes. Examples can be seen in figure 6.8. Image 6.8a contains fibrous tissue as well as acinar adenocarcinoma, image 6.8b contains in the upper left corner cartilage, in the bottom right one glandular tissue and image 6.8c contains in the bottom left corner normal alveolar lung parenchyma that changes in the upper right corner to papillary adenocarcinoma. All of these cases have to be referred to one particular class to define the ground truth. If the learning algorithm classifies these tiles into the other class that the one of the ground truth, this is treated as an error, but in truth it is not completely wrong. This have to be kept in mind during the determination of the classification accuracy. The values can only give

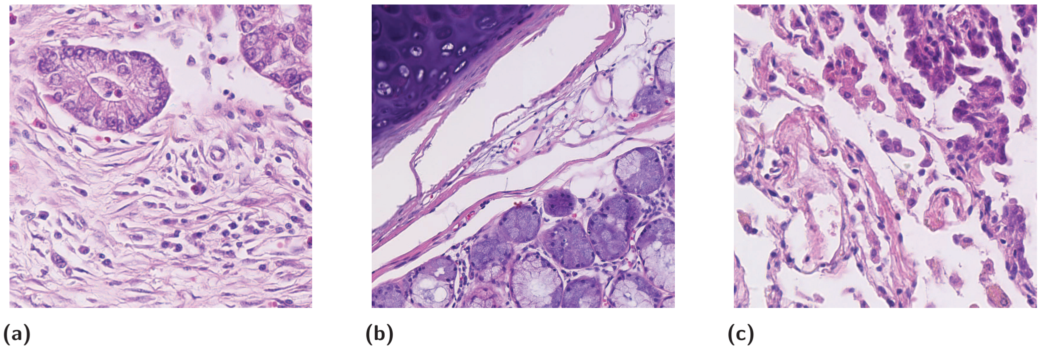


Figure 6.8: Three tiles with tissue areas that contain different classes.

an indication of the accuracy.

6.2.2 Classification Results

The classification accuracy for the 9 images with available ground truth with the feature subset 1 is 67.16 %. With the feature set containing wavelet features a slightly better accuracy of 68.01 % could be achieved. If just the accuracy for the tumor classes is observed, the classification with subset 1 achieves 61.26 %, with subset 2 64.27 %. More detailed classification results can be seen in tables 6.2 and 6.3, in which the confusion matrices for the classification with the two feature sets are shown. One can see that the classes normal alveolar lung parenchyma and lepidic adenocarcinoma have for both feature sets a very high true positive rate. On the other hand normal bronchi and micropapillary have very low true positive rates. This can be explained with the very low number of example tiles. Furthermore one can see that some classes are commonly mislabeled to another class. Macrophage infiltrates for example are often predicted as normal alveolar lung parenchyma or fibrosis, most likely because macrophages often occur in combination with these classes. A high misclassification rate can also be seen with papillary adenocarcinoma. With feature subset 1 a lot of tiles are wrongly classified as normal alveolar lung parenchyma, acinar- or mucious adenocarcinoma. At least the wrong classification to mucious is reduced by including wavelet features. Also the other case, that tiles that are in truth mucious but are labeled as papillary, could be reduced.

	true normal alveolar lung parenchyma	true acinar adeno carcinoma	true mucious adeno carcinoma	true papillary adeno carcinoma	true micropapillary adeno c.	true solid adeno carcinoma	true lepidic adeno carcinoma	true cartilage	true normal bronchi	true glandular tissue	true fibrosis	true macrophage infiltrates	true blood vessels	true lymphocytes	class precision
pred. normal alveolar lung parenchyma	17001	308	186	1856	104	229	52	50	357	104	1878	619	674	42	72,47%
pred. acinar adeno carcinoma	161	5459	165	1651	37	678	3	34	38	43	589	74	36	59	60,47%
pred. mucious adeno carcinoma	47	206	2809	1603	180	282	6	27	58	184	442	60	35	27	47,08%
pred. papillary adeno carcinoma	121	410	192	2521	21	61	55	26	20	41	261	45	37	19	65,82%
pred. micropapillary adeno carcinoma	0	0	1	0	22	0	0	0	0	0	0	0	0	0	95,65%
pred. solid adeno carcinoma	9	158	47	73	0	4816	1	0	12	7	332	30	11	100	86,06%
pred. lepidic adeno carcinoma	125	2	0	1	0	0	3872	10	21	0	92	63	65	0	91,08%
pred. cartilage	5	2	0	0	0	0	0	299	0	2	3	0	0	0	96,14%
pred. normal bronchi	0	0	1	0	0	0	0	1	5	1	1	0	0	0	55,56%
pred. glandular tissue	1	1	0	0	0	0	0	5	2	252	11	0	0	0	92,65%
pred. fibrosis	645	1050	354	708	5	1257	7	26	123	91	6578	501	391	116	55,50%
pred. macrophage infiltrates	357	8	1	12	2	2	0	0	3	0	146	677	6	1	55,72%
pred. blood vessels	25	1	1	2	0	2	5	1	7	1	59	1	478	0	81,99%
pred. lymphocytes	0	5	0	0	0	338	0	0	0	0	94	0	3	331	42,93%
class recall	91,91%	71,73%	74,77%	29,92%	5,93%	62,83%	96,78%	62,42%	0,77%	34,71%	62,73%	32,71%	27,53%	47,63%	

Table 6.2: Confusion matrix of classification results with feature subset 1.

The classification accuracies for each of the whole slide images separately can be seen in figure 6.4. The best result for both feature sets is achieved for image no. 9. The result achieved with feature subset 2 can be seen in figure 6.9. Normal alveolar lung parenchyma (green) is detected by the classification algorithm as well as lepidic adenocarcinoma (violet) which is the dominant subtype in this tumor section.
The accuracy with feature subset 2 of image no. 5 is over 10 % lower than the one of image no. 9, but in figure 6.10 one can see that the result is still very good. Again, the most tiles containing normal alveolar lung parenchyma are classified correctly by the algorithm. Furthermore the detected predominant adenocarcinoma subtype is papillary (orange). This corresponds to the ground truth.
For the images no. 3 and no.7 the worst classification accuracies are achieved. These are the two image from which no tiles are used for the generation of the training sample set, which is one explanation of the low accuracy values.
The highest difference in the accuracy occurs for image no. 3. The classification results with both feature sets and the ground truth for a part of this image is shown in figure 6.11. The main difference is that with feature subset 1 big areas that belong in truth to the papillary subtype (orange) are classified as mucious (pink).

An important aspect for the feature set used to classify the lung tumor sections is the calculation time. One factor for this is the number of used features. In table 6.6 all features of feature subset 1 and 2 are listed. In the notation of the feature vectors channel 1,2,3 name the RGB color channels and 3,4,5 the HSV color channels respectively. The first feature set without wavelet features contains 72 features, the second set with wavelet features 96. Furthermore it stands out that the chosen features are very equal in both cases. For example the intensity attributes on channel 2 and magnification 6 are omitted in both sets.

To build the feature sets 11 features vectors have to be calculated in both cases. Nonetheless one can see in figure 6.5, in which the calculation time for 3 images is shown, that the calculation time of feature subset 1, which does not contain wavelet features, is considerably shorter.

	true normal alveolar lung parenchyma	true acinar adeno carcinoma	true mucious adeno carcinoma	true papillary adeno carcinoma	true micropapillary adeno c.	true solid adeno carcinoma	true lepidic adeno carcinoma	true cartilage	true normal bronchi	true glandular tissue	true fibrosis	true macrophage infiltrates	true blood vessels	true lymphocytes	class precision
pred. normal alveolar lung parenchyma	16970	266	162	1625	87	224	82	30	362	130	1858	642	716	31	73,19%
pred. acinar adeno carcinoma	138	5540	125	1778	43	580	2	11	29	62	499	84	22	51	61,80%
pred. mucious adeno carcinoma	85	179	3002	792	192	182	3	61	33	190	355	31	45	27	57,99%
pred. papillary adeno carcinoma	216	450	89	3104	35	97	30	2	20	84	457	125	48	24	64,92%
pred. micropapillary adeno carcinoma	0	0	0	1	9	0	0	0	0	0	0	0	0	0	90,00%
pred. solid adeno carcinoma	9	191	57	418	0	4926	1	0	8	7	395	42	15	107	79,76%
pred. lepidic adeno carcinoma	111	2	0	0	0	0	3877	4	23	0	83	72	57	0	91,68%
pred. cartilage	8	0	0	0	0	2	0	346	1	2	7	0	0	1	94,28%
pred. normal bronchi	0	0	0	0	0	0	0	0	4	1	4	0	0	0	44,44%
pred. glandular tissue	2	0	0	0	0	0	0	0	1	112	8	0	1	0	90,32%
pred. fibrosis	687	962	321	694	3	1320	6	25	158	136	6601	574	456	136	54,65%
pred. macrophage infiltrates	254	16	0	13	2	2	0	0	4	0	108	500	2	1	55,43%
pred. blood vessels	17	0	1	1	0	0	0	0	2	2	15	0	372	0	90,73%
pred. lymphocytes	0	4	0	1	0	332	0	0	1	0	96	0	2	317	42,10%
class recall	91,74%	72,80%	79,90%	36,83%	2,43%	64,27%	96,90%	72,23%	0,62%	15,43%	62,95%	24,15%	21,43%	45,61%	

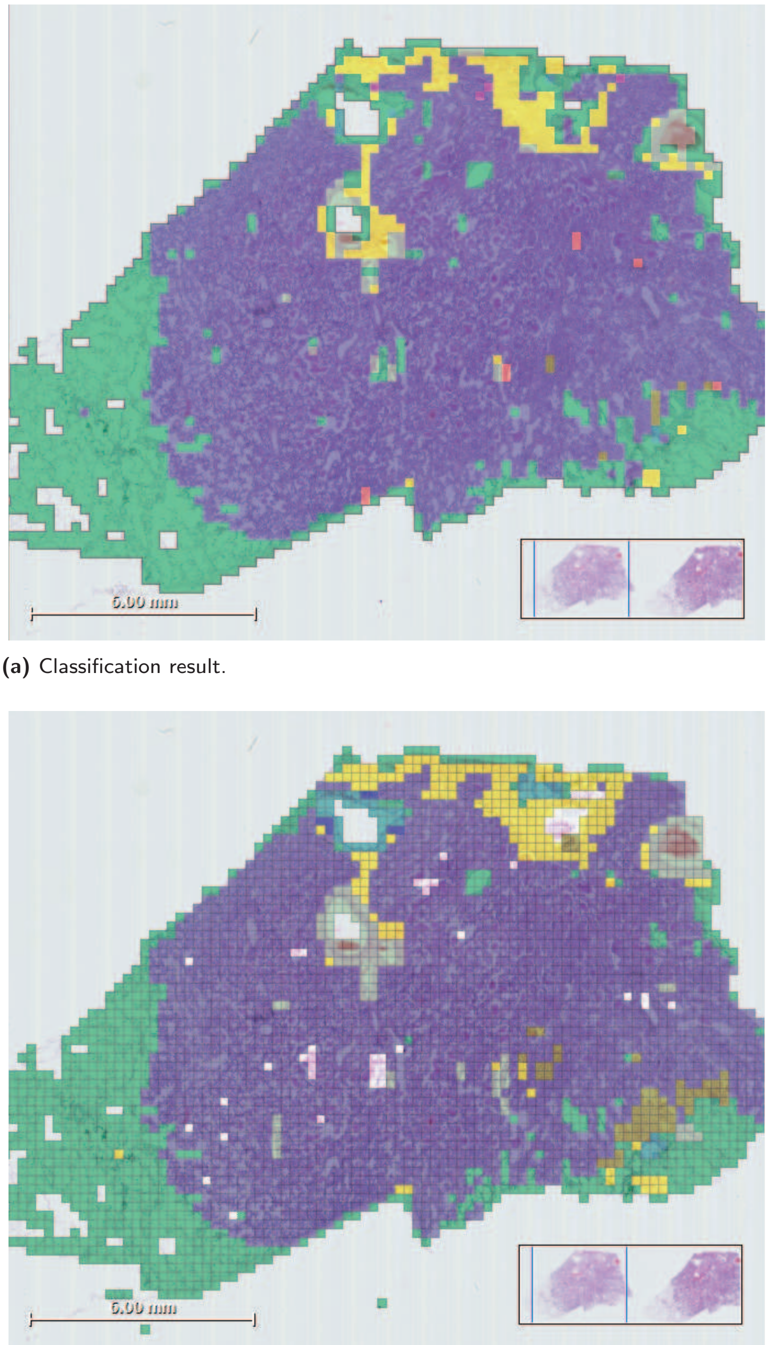
Table 6.3: Confusion matrix of classification results with feature subset 2 which contains wavelet features.

image no.	accuracy with subset 1	accuracy with subset 2
	(without wavelet features)	(with wavelet features)
1	74 %	76 %
2	67 %	65 %
3	46 %	50 %
4	79 %	81 %
5	77 %	73 %
6	73 %	70 %
7	52 %	54 %
8	60 %	61 %
9	88 %	87 %

Table 6.4: Comparison of the classification accuracy of the subset including wavelet features with the one without wavelets for each image with available ground truth.

image.	number of	calc. time subset 1	calc. time subset 2
	classified tiles	(without wavelet features)	(with wavelet features)
a	6935	1:14	1:27
b	8878	1:23	2:17
c	15796	3:25	4:43

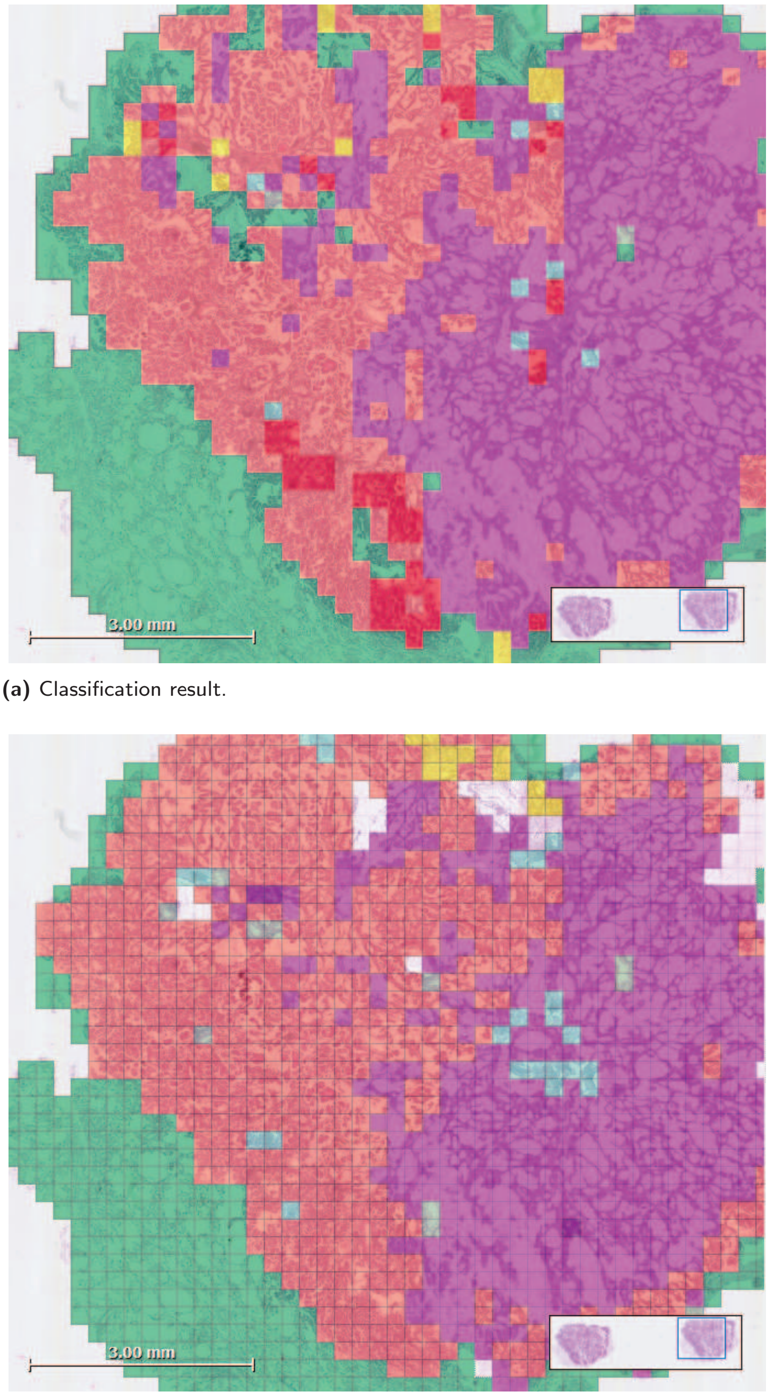
Table 6.5: Calculation time (in min:sec) for 3 histological images of different size.



(a) Classification result.

(b) Ground truth.

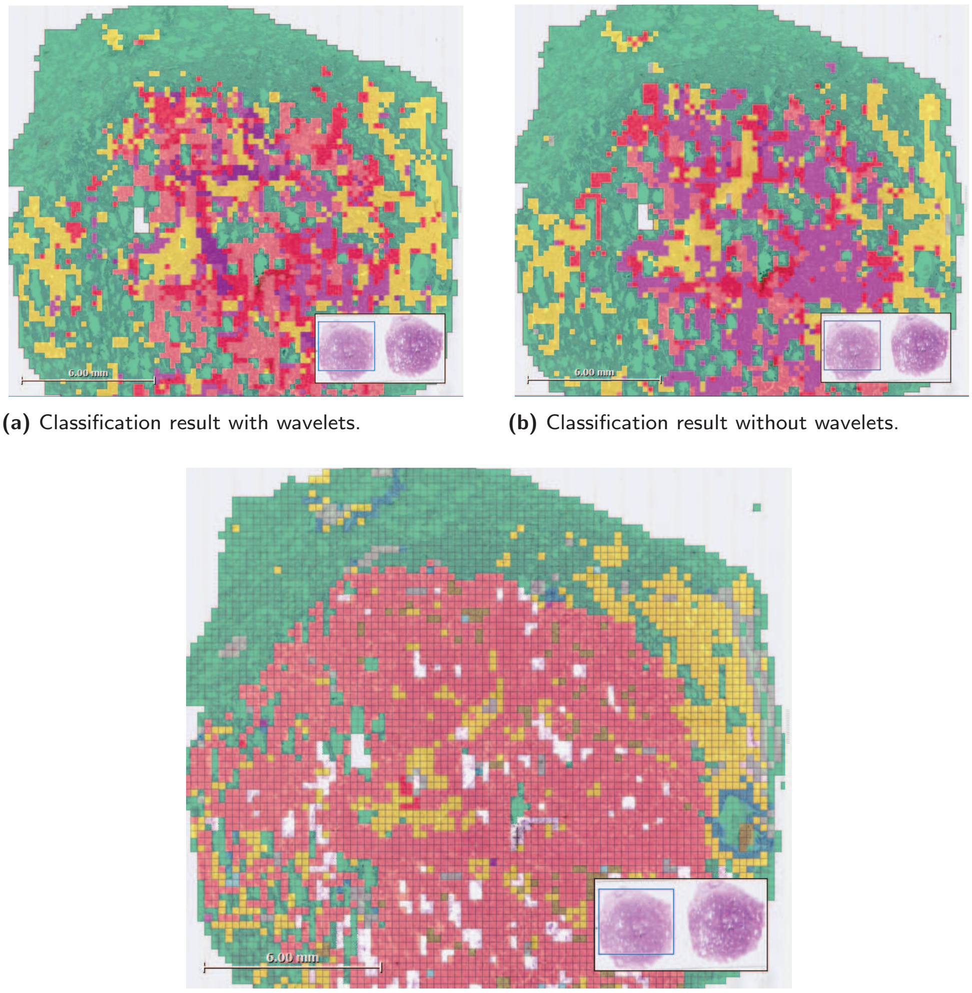
Figure 6.9: Part of the classification result with HistoCAD of image no. 9.



(a) Classification result.

(b) Ground truth.

Figure 6.10: Part of the classification result with HistoCAD of image no. 5.



(a) Classification result with wavelets.

(b) Classification result without wavelets.

(c) Ground truth.

Figure 6.11: Part of the classification result with HistoCAD of image no. 3.

feature vector	chosen in subset 1 (without wavelet features)	chosen in subset 2 (with wavelet features)
mag-6-channel-0-intensity-attributes	min, max, mean, stddev, lower quartile	min, max, mean, lower quartile, upper quartile
mag-6-channel-3-intensity-attributes	min, stddev, median, upper quartile	min, median, upper quartile
mag-12-channel-0-intensity-attributes	max, stddev, median, upper quartile	max, lower quartile, median, upper quartile
mag-25-channel-0-intensity-attributes	max, stddev, median, lower- , upper quartile	max, stddev, median, lower- , upper quartile
mag-25-channel-3-intensity-attributes	mean, stddev, lower quartile	—-
mag-12-channel-0-lbp-attributes	pattern #3, #4, #5, #7, #8	pattern #3, #4, #5, #7, #8
mag-12-channel-1-lbp-attributes	pattern #4, #5, #7, #8	pattern #3, #4, #5, #7, #8
mag-12-channel-4-lbp-attributes	pattern #3, #4, #5	—-
mag-25-channel-1-lbp-attributes	pattern #1, #5, #6, #7	pattern #1, #4, #5, #7
mag-50-channel-0-lbp-attributes	pattern #2, #3, #5	—-
mag-50-channel-1-lbp-attributes	pattern #1, #5, #8	—-
mag-50-channel-3-lbp-attributes	—-	pattern #4, #5, #7, #8
mag-6-channel-3-wavelet-attributes	—-	$f_1(1), f_1(2), f_1(3)$
mag-25-channel-3-wavelet-attributes	—-	$f_1(3), f_1(4), f_1(5),$ $f_2(4), f_2(6)$
mag-50-channel-3-wavelet-attributes	—-	$f_1(1), f_1(2), f_1(3), f_1(4),$ $f_1(5), f_2(3), f_2(4), f_2(5)$

Table 6.6: Selected features for the feature subsets 1 and 2.

7 Chapter 7 Conclusion and Outlook

The aim of this work was to classify tissue sections of lung adenocarcinoma automatically and it was evaluated if the classification can be improved by including wavelet based image features.

The first goal was to achieve a good differentiation between tumor tissue and healthy lung tissue. The classification results show that this is possible. The presented methods achieve true positive rates of over 90% for normal alveolar lung parenchyma. This tissue class comprises the greatest part of healthy lung tissue and should therefore not falsely be classified as tumor. With this high detection rate it can be distinguished to lung tumor tissue.
However, the classification of lung adenocarcinoma into its different subtypes is still a challenge. The mean accuracy among all histological images is approximately 67 %. The mean accuracy just for the adenocarcinoma subtypes is 61.26 % for the classification without and 64.27 % with wavelet features. This is insufficient for a reliable automatic classification. For lepidic adenocarcinoma very good classification results could be achieved, whereas the results for the papillary subtype contain still a lot of misclassifications. This shows that a complete automatic classification is, at this point, impossible. The task is too complex.

For an improvement of the classification an additional texture feature based on wavelets was introduced in this work. Although better classification results for some tumor classes were achieved by including these features, the overall accuracy is not significantly improved. The local binary patterns, which were already available for other classification tasks, are already a good simple texture feature. Additionally including wavelet based features is in this application not profitable, because the considerably increased calculation time does not go along with an adequate improvement of the classification accuracy.

Probably, a sufficient classification accuracy of the lung adenocarcinoma subtypes could be achieved with a semi-automatic classification. In this scenario the classification algorithm pre-classifies the data. Afterwards a pathologist could correct falsely classified tiles. These tiles are added to the training data set and the classification is repeated. This semi-automatic approach has to main advantages. First, the training data set could interactively be improved by adding new characteristic tiles of a tumor class or tiles that are difficult to classify, which were possibly not available in the first training data set. Secondly, this method would still help the pathologist in the quantization of the tumor types. The pathologist does not have to determine the occurrence of a tumor type by hand because this is already given with the classification result.

Furthermore, a better classification result could be achieved by taking different tile sizes into account. With a smaller tile size class borders could be captured more clearly and the error in tiles which include different classes could be reduced. But this has to be done by considering the calculation time. One idea is to calculate the feature on the current tile size while the classification is done on a finer grid.

Bibliography

[1] *Krebs in Deutschland 2005/2006. Häufigkeiten und Trends*. Robert Koch Institut und Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V., 2010.

[2] BREIMAN, L. Random forests. *Machine Learning 45* (2001), 5–32.

[3] BURRUS, C. S., GOPINATH, R., AND GUO, H. *Introduction to Wavelets and Wavelet Transforms: A Primer*. Prentice Hall, 1997.

[4] CHANG, T., AND KUO, C. Texture analysis and classification with tree-structures wavelet transform. *IEEE Transactions on Image Processing 2*, 4 (1993), 429–441.

[5] COHEN, A., DAUBECHIES, I., AND FEAUVEAU, J. Biorthogonal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics 45* (1992), 465–560.

[6] DAUBECHIES, I. Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics 41* (1988), 909–996.

[7] DAUBECHIES, I. *Ten Lectures on Wavelets*. No. 61 in CBMS/NSF Series in Applied Math. Society for Industrial and Applied Mathematics (SIAM), 1988.

[8] HALL, M. *Correlation-based Feature Selection*. PhD thesis, University of Waikato, 1998.

[9] HALL, M., FRANK, E., HOLMES, G., PFAHRINGER, B., REUTEMANN, P., AND WITTEN, I. The weka data mining software: an update. *SIGKDD Explorations 11*, 1 (2009), 10–18.

[10] HOOGSTRATEN, B., ADDIS, B., HANSEN, H., MARTINI, N., AND SPIRO, S. *Lung Tumors: Lung, Mediastinum, Pleura and Chest Wall*. Current Treatment of Cancer. Springer, 1988.

[11] J. SOBOTTA, U. W. *Atlas Histologie: Zytologie, Histologie, Mikroskopische Anatomie*, 7 ed. Urban & Fischer Verlag/Elsvier GmbH, 2005.

[12] LESSMANN, B., NATTKEMPER, T., HANS, V., AND DEGENHARD, A. A method for linking computed image features to histological semantics in neuropathology. *Journal of Biomedical Informatics 40* (2007), 631–641.

[13] LIVENS, S., SCHAUNDERS, P., VAN DE WOUWER, G., VAN DYCK, D., SMETS, H., WINKELMANS, J., AND BOLGAERTS, W. A texture analysis approach to corrosion image classification. *Microscopy, Microanalysis, Microstrucures 7*, 2 (1996), 1–10.

[14] LIVENS, S., SCHEUNDERS, P., VAN DE WOUWER, G., AND VAN DYCK, D. Wavelets for texture analysis, an overview. In *6th International Conference on Image Processing and Its Applications* (1997), vol. 2, pp. 581–585.

[15] LÜLLMANN-RAUCH, R. *Taschenlehrbuch Histologie*. Georg Thieme Verlag, 2009.

[16] MA, W., AND MANJUNATH, B. A comparison of wavelet transform features for texture image annotation. In *ICIP '95: Proceedings of the 1995 International Conference on Image Processing* (1995), IEEE Computer Society, pp. 256–259.

[17] MALLAT, S. A theory for multiresolution signal decomposition: The wavelet represenation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 11*, 7 (1989), 674–693.

[18] MIERSWA, I., WURST, M., KLINKENBERG, R., SCHOLZ, M., AND EULER, T. YALE: Rapid prototyping for complex data mining tasks. In *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining* (2006), ACM, pp. 935–940.

[19] OJALA, T., AND PIETIKÄINEN, M. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence 24*, 7 (2002), 971 – 987.

[20] RADEN, T., AND HUSOY, J. Filtering for texture classification: A comparative study. *IEEE Transactions on Pattern Analysis and Machine Intelligence 21*, 4 (1999), 291–310.

[21] SMITH, J. R., AND CHANG, S. Transform features for texture classification and discrimination in large image databases. In *ICIP '94: Proceedings of the 1994 International Conference on Image Processing* (1994), IEEE Computer Society, pp. 407–411.

[22] STRANG, G., AND NGUYEN, T. *Wavelets and Filterbanks*. Wellesly-Cambridge Press, 1996.

[23] TRAVIS, W. D., BRAMBILLA, E., AND NOGUCHI, M. IASLC/ATS/ERS international multidisciplinary classification of lung adenocarcinoma. *Journal of Thoracic Oncology 6*, 2 (2011), 244–285.

[24] WITTEN, I. H., AND FRANK, E. *Data Mining: Practical Machine Learning Tools and Techniques*, 2 ed. Elsvier, 2005.