# Blood Vessel Classification into Arteries and Veins in Retinal Images

Claudia Kondermann and Daniel Kondermann[a] and Michelle Yan[b]

[a]Interdisciplinary Center for Scientific Computing (IWR), University of Heidelberg , Im Neuenheimer Feld 368, 69120 Heidelberg, Germany

[b]Michelle Yan, Siemens Corporate Research, 755 College Road East, Princeton, NJ, 08540, USA

## ABSTRACT

The prevalence of diabetes is expected to increase dramatically in coming years; already today it accounts for a major proportion of the health care budget in many countries. Diabetic Retinopathy (DR), a micro vascular complication very often seen in diabetes patients, is the most common cause of visual loss in working age population of developed countries today. Since the possibility of slowing or even stopping the progress of this disease depends on the early detection of DR, an automatic analysis of fundus images would be of great help to the ophthalmologist due to the small size of the symptoms and the large number of patients. An important symptom for DR are abnormally wide veins leading to an unusually low ratio of the average diameter of arteries to veins (AVR). There are also other diseases like high blood pressure or diseases of the pancreas with one symptom being an abnormal AVR value. To determine it, a classification of vessels as arteries or veins is indispensable. As to our knowledge despite the importance there have only been two approaches to vessel classification yet. Therefore we propose an improved method. We compare two feature extraction methods and two classification methods based on support vector machines and neural networks. Given a hand-segmentation of vessels our approach achieves 95.32% correctly classified vessel pixels. This value decreases by 10% on average, if the result of a segmentation algorithm is used as basis for the classification.

**Keywords:** Vessel Classification, Arteries and Veins, Pattern Recognition, Neural Networks, Support Vector Machines, Diabetic Retinopathy

## 1. INTRODUCTION

There are two types of vessels, arteries and veins. Arteries are brighter, since they transport blood rich in oxygen to the organs of the body. The veins afterwards transport the blood, which is at a low oxygen level and thus darker, to the lungs and the liver. For many medical applications it would be of great benefit, if the vessels could be distinguished into arteries and veins, since there are many diseases with one symptom being an abnormal ratio of the size of arteries to veins. For example in diabetic patients the veins are abnormally wide, while diseases of the pancreas lead to narrowed arteries and high blood pressure results in thickened arteries. To detect these diseases the retina is routinely examined.

As a basis for classification a good segmentation of blood vessels is of course needed (see [Leandro2001][1] , [Gang2002][2] or [Hoover2000][3]). There are mainly four different features that can be used to distinguish arteries from veins in general:

- arteries are brighter in color than veins
- arteries are thinner than neighboring veins
- the central reflex (the light reflex of the inner parts of the vessels shown in Figure 1) is wider in arteries and smaller in veins.
- arteries and veins usually alternate near the optic disk before branching out; that means near the optic disk one artery is usually next to two veins and the other way round

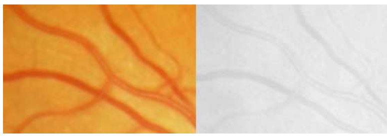


**Figure 1.** One of the most important features for the discrimination of arteries and veins is the central reflex in the red channel; left: original image containing two large veins and an artery in the center, right: red channel

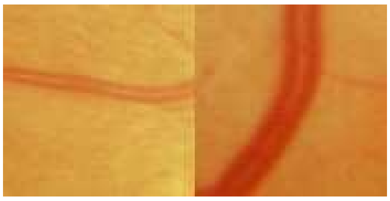


**Figure 2.** Typical representative of arteries (left) and veins (right), they can be discriminated by color, size, central reflex size and topological properties

One typical representative of each arteries and veins can be seen in Figure 2.

The mentioned features often provide enough information to successfully classify a vessel as artery or vein. However, in many cases they do not suffice for the following reasons:

- If the image quality is not good enough - which is especially the case in the outer regions of the image - the central reflex often vanishes.
- Vessels in the outer regions of the image are very dark due to the shading effect (inhomogeneous lighting of the image). Here arteries and veins look very much alike, which necessarily leads to the misclassification of some vessels.
- The width of the vessel is also not very useful for classification, since it changes being largest near the optic disk and smallest on the outer parts of the image.
- The alternation of arteries and veins only holds true for the vessels very near to the optic disk. When they start branching out, it is common that two branches of the same vessel lie next to each other.

So none of the typical features of arteries and veins is globally valid. Figure 3 shows examples for all four problems.

To give an overall impression of the difficulty of this classification task, ten cropped veins and ten cropped arteries taken from four different retinal images can be seen in Figure 4. The quality of the images, the background and the small size of the vessels and the subtleness of the features themselves make it very hard to distinguish between the two classes. These examples make clear that a classification method based only on local features will not be able to achieve good results. We combine these features in a learning based approach, which - with the help of global meta-knowledge - is able to distinguish arteries from veins with a very high classification rate.

## 2. RELATED WORK

There have not been many approaches to the vessel classification problem yet. In [Li2003][4] a piecewise Gaussian model especially adapted to the central reflex of the vessels is used as a filter on the image for classification. The idea in [Grisan2003][5] is to divide the retinal image into four quadrants within a concentric zone around the optic disk in order to handle the discrimination of the vessels locally in each quadrant. For each quadrant separately the variance of the red values and the mean of the hue values of the pixels is used as a feature vector.

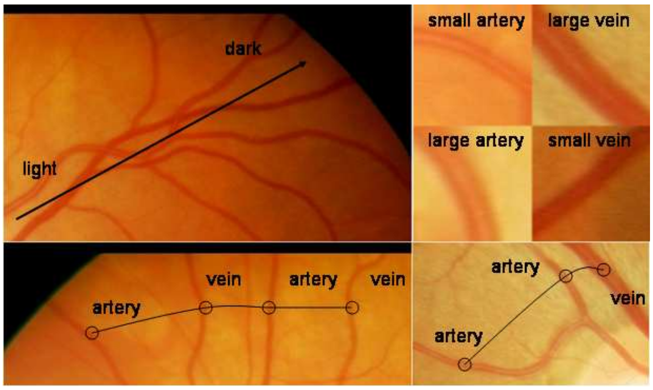



**Figure 3.** top left: In the outer regions of the image the vessels are darker and the central reflex is not visible anymore, which makes classification very hard; top right: The size of the vessel cannot be used as global discrimination feature; bottom left: Example of four larger vessels, where discrimination into arteries and veins is very difficult; bottom right: The alternation of arteries and veins is not valid because of artery branch

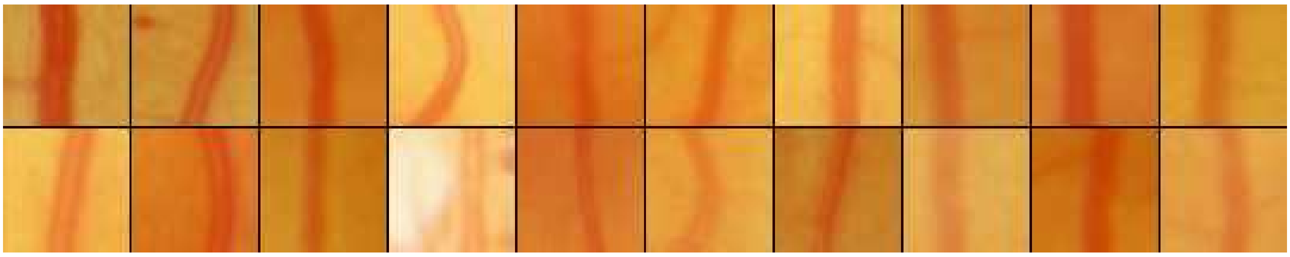


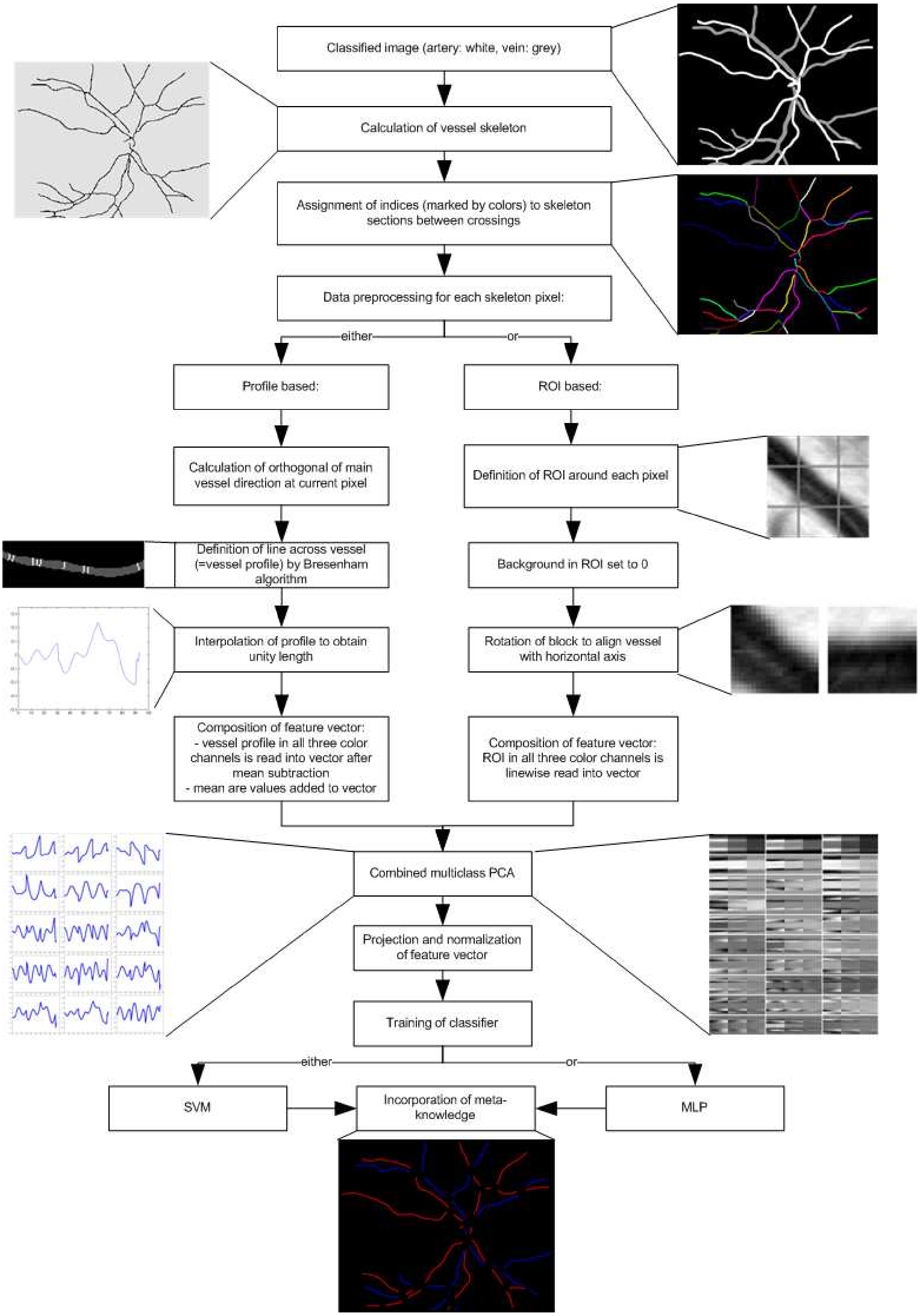
**Figure 4.** Different examples for arteries and veins illustrating the difficulty of the classification task, first row: ten different cropped veins, second row: ten different cropped arteries, all taken from four different fundus images

## 3. OVERVIEW

We first have to preprocess the image in order to remove the shading effect to at least some extent. Then we develop and compare two different methods to create feature vectors for the classification task. One is based on the vessel profile, the other on the region of interest (ROI) of a centered vessel pixel. In this way we obtain a good representation of the central reflex and the color of the vessel. To reduce the dimensionality of the feature vectors, we apply combined multiclass PCA to the resulting feature vectors as described in [Nieuwenhuis2006][6] . This data can then be used to train and apply a SVM and a neural network. In order to train both classifiers several images with segmented vessels classified as arteries and veins are necessary. The classifier can then be applied to any image with already segmented vessels. To diminish the chance of misclassification due to insufficient local features we finally infer some meta-knowledge about vessels by making use of the fact that a vessel does not change its class between two crossings. An overview of the classification algorithm is presented in Figure 3.

## 4. PREPROCESSING

To improve the results, three preprocessing steps are necessary: image enhancement, extraction of the skeleton of the segmented vessel tree and identification of connected vessel sections between each two crossings. 1) For vessel classification we use all three color channels of the original image. Here the shading effect poses a major problem as it is responsible for large darker regions in the image. These regions firstly make it hard to even recognize a vessel against the background and secondly lead to misclassifications of vessels, because of their very dark color and the missing central reflexes in these regions. To reduce this problem we approximate the background of the image as a 2-dimensional spline and subtract this spline from the image. This is done for each of the color channels separately.


Classified image (artery: white, vein: grey)
Calculation of vessel skeleton
Assignment of indices (marked by colors) to skeleton sections between crossings
Data preprocessing for each skeleton pixel:
either
or
Profile based:
ROI based:
Calculation of orthogonal of main vessel direction at current pixel
Definition of ROI around each pixel
Definition of line across vessel (=vessel profile) by Bresenham algorithm
Background in ROI set to 0
Interpolation of profile to obtain unity length
Rotation of block to align vessel with horizontal axis
Composition of feature vector:
- vessel profile in all three color channels is read into vector after mean subtraction
- mean are values added to vector
Composition of feature vector:
ROI in all three color channels is linewise read into vector
Combined multiclass PCA
Projection and normalization of feature vector
Training of classifier
either
or
SVM
Incorporation of meta-knowledge
MLP

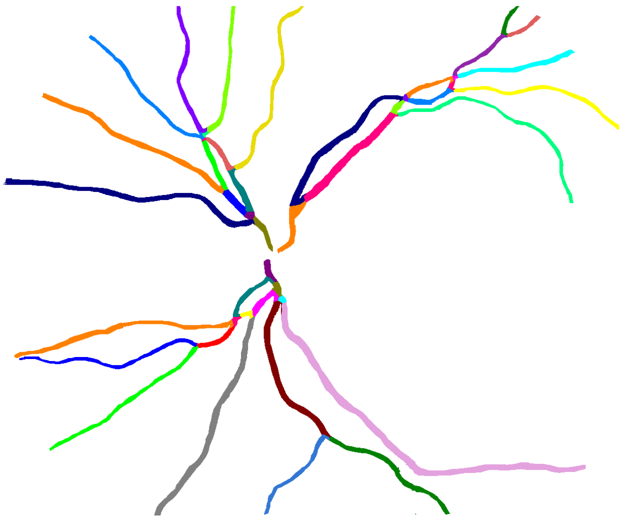


**Figure 5.** Vessel tree partitioned into sections between crossings

2) The skeleton of the segmented vessel tree is extracted in order to reduce the complexity of the classification, since it is sufficient to classify only the pixels on the centerlines of the vessels.
3) As we know that pixels between two crossings in the vessel tree must belong to the same vessel class, we identify the crossings in the vessel skeleton and assign a unique section number to all pixels between each two crossings. An example can be seen in Figure 5.

We now turn to the issue of extracting feature vectors from the image which can then be used as training data. We compare two different feature vector extraction methods: profile based and ROI based.

## 5. FEATURE VECTOR EXTRACTION

We compare two methods to extract feature vectors for classification: one is based on the vessel profile the other on a region of interest around the vessel pixel on the center line.

### 5.1. Vessel Profile Based Feature Vectors

To create the vessel profile based training data we read the profile of the vessel into a feature vector. The main idea is that for each skeleton pixel we determine the orthogonal direction to the general direction of the vessel at the current center line position, then draw a line across the vessel and read the vessel profile along this line into a vector. We achieve the best results, if the profile consists of color information of all three color channels after subtraction of the mean value, which is added at the end of each color profile. As the vessels show different widths leading to different lengths of the profile feature vectors, we use spline interpolation to obtain unity length for each feature vector.

### 5.2. Vessel ROI Based Feature Vectors

We now demonstrate how to obtain vessel ROI based feature vectors as training data. The idea is to define a quadratic region of interest around each skeleton pixel. In order to reduce the difficulty of the learning task for the classifier, the ROI is centered on the center of gravity of the vessel inside the ROI and rotated to align the main axis of the vessel with the horizontal axis. Afterwards all three color channels of the region are read linewisely into the feature vector. Results of the rotation can be seen in Figure 6.

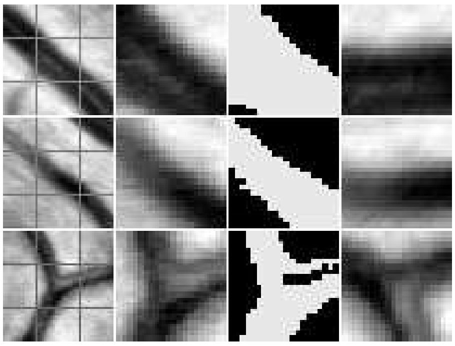


**Figure 6.** Each row contains a rotation example, first column: center square is ROI chosen for rotation, second column: ROI before rotation, third column: binarized ROI, fourth column: ROI after rotation

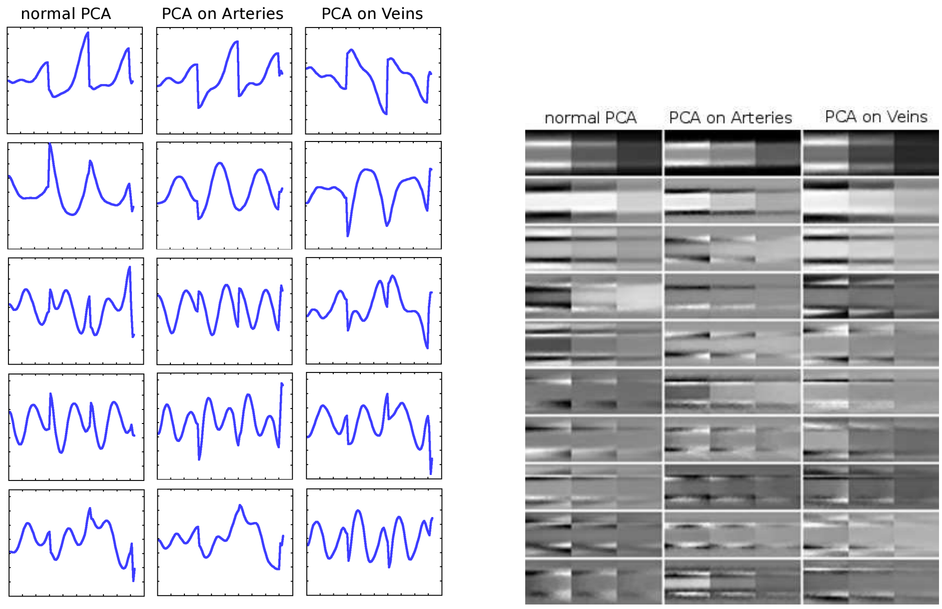



**Figure 7.** left: first five eigenprofiles based on the profile feature vector composition method, right: resulting first ten eigenvessels based on the ROI feature vector composition method

## 6. COMBINED MULTICLASS PCA

To reduce the dimension of the training data we now apply combined multiclass PCA [Nieuwenhuis2006][6] to the feature vector. To define the feature classes for the combined multiclass PCA we specify arteries as one class and veins as the other class. The resulting set of principal components then consists of three sections, that is the components derived from PCA carried out on 1) all training data samples, 2) artery data samples only, 3) vein data samples only. For both feature vector composition methods the resulting first principal components of all three classes can be seen in Figure 7.

For the ROI based feature vectors, differences between the eigenvectors of all three classes are obvious on the right in Figure 7. Among the artery eigenvessels for example the second eigenvessel indicates the generally smaller size and the thinner edges of arteries when compared to the second eigenvessels of the other two classes. The second eigenvector resulting from normal PCA seems to be an average of the second vectors of arteries and veins.

For the profile based feature vectors the results confirm the following: One of the most important features for the discrimination of arteries and veins is the central reflex in the red channel. Veins show larger central reflexes and vessel edges as well as larger color differences in the red channel between edge and reflex (see Figure 1). Contrary, arteries are lighter then veins. These features can be found in the second eigenprofiles of arteries and veins on the left in Figure 7.

## 7. CLASSIFICATION

We now present two classification methods for the feature vectors after dimensionality reduction: support vector machines and neural networks. Both classification methods can be combined with both presented feature vector composition methods.

### 7.1. Support Vector Machines

As a first classification method we use a support vector machine (SVM). We have compared the classification results of four kernels: linear, polynomial, RBF and sigmoidal with various parameters. The best result yielded the following RBF kernel

$$k(\mathbf{x_i}, \mathbf{x_j}) = (-0.0625 \cdot ||\mathbf{x_i} - \mathbf{x_j}||^2) + 4096 \tag{1}$$

### 7.2. Neural Networks

As a second classification method we use a neural network, a multilayer perceptron (MLP), trained with the backpropagation algorithm. The input layer size depends on the chosen feature vector composition method. We compared one and two hidden layers with various hidden layer sizes between five and fifty neurons and obtained the best result for forty hidden neurons in one layer, that is for the size combination (60-40-2) for input, one hidden and output layer.

## 8. INCLUDING META-KNOWLEDGE TO IMPROVE THE CLASSIFICATION RESULT

After classification each pixel has been assigned to either the artery or the vein class. To improve the result we can infer some meta-knowledge. The concept is that all pixels belonging to the same vessel section between two crossings must also belong to the same class. To use this knowledge we now need the skeleton sections, that have been calculated during the preprocessing step. We add up the confidence of all pixels in one section for belonging to the artery and for belonging to the vein class. Then we choose the class label with the higher value. In this way the pixels classified with high reliability will influence the final classification of the whole section to a larger extent than the ones with poor reliability. This method is especially helpful for longer vessel sections that start somewhere near the center and end in the outer regions of the image, since in this case the pixels far away from the center are usually very hard to classify correctly. If we use a neural network we cannot make use of direct confidence information. But we can use the answer of the two output neurons $(a, v)$ for arteries and veins to produce a confidence measure as

$$p(a, v) \begin{cases} 1 - \sqrt{(1-a)^2 + v^2} \text{ if } a \geq v \\ 1 - \sqrt{a^2 + (1-v)^2} \text{ otherwise} \end{cases} \tag{2}$$

An example for the improvement of a classification result by the use of meta-knowledge can be seen in Figure 8.

## 9. RESULTS

Two feature vector composition methods - profile based and ROI based - have been presented together with two classification methods, SVMs and neural networks. Since only the main vessels near the optic disk are clinically relevant, we present results for the classification of the vessel pixels within three diameters of the optic disk. To assess the performance of each combination of a feature vector composition method and a classifier we have conducted tests on four 1024x1280 retina images containing 10132 different vessel skeleton pixels. The images are of high quality, and the optic disk is in the center of the image so the main vessels are not so much affected by the shading effect. The results of each feature vector composition method combined with each classification method can be found in Table 1. An example for a whole classified image can be seen in Figure 9.

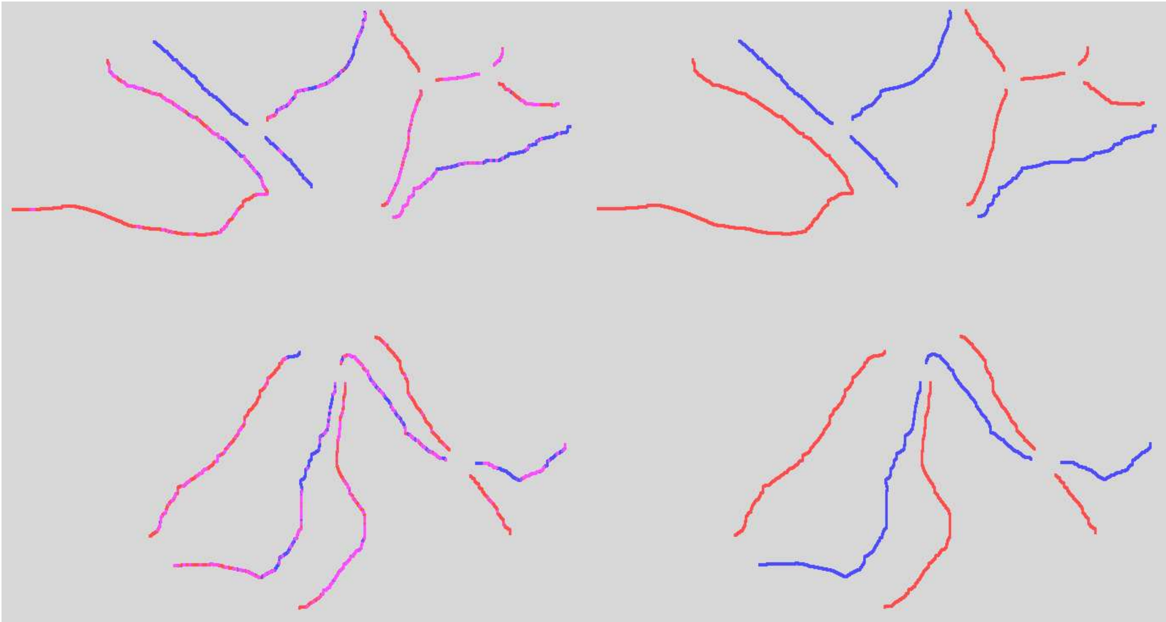


**Figure 8.** Results of application of meta-knowledge after classification, left: classification result, right: result after meta-knowledge application (red: arteries, blue: veins)

| | **Best Classification Results** | |
|---|---|---|
| | Neural Net | SVM |
| Profile Based | 94.39% | 93.44% |
| ROI Based | 95.32% | 92.84% |

**Table 1.** Comparison of the best results of each combination of feature vector and classifier; percentage of correctly classified vessel skeleton pixels

There are several conclusions we can draw from the results. First, the differences in the results of all four combinations are minor, so all methods yield very good results. The best combination is the ROI based feature vector and the neural net classifier obtaining 95.32% correctly classified pixels. The inclusion of meta-knowledge increases the number of correct classifications by 6.44%. The cases in which the results deteriorate after including meta-knowledge are very rare and usually confined to the rather poor results. The reason for this is obvious, since in good results already most of the pixels in a vessel have been classified correctly and there are only few misclassifications which can easily be corrected by the majority of the pixels in the vessel. In contrast to that if the result is poor the majority of the pixels in the vessel has been misclassified, so the use of meta-knowledge even increases these numbers. The quality of the images differs a lot and has a large impact on the results of the classification algorithm. If we use the output of a segmentation algorithm instead of hand-segmented images as basis for the classification the results deteriorate by 10% on average.

## 10. SUMMARY AND CONCLUSION

We have presented two methods for feature extraction and two methods for classification. Interesting results are that neural networks and SVMs are about equally well suited for the described classification task and achieve extremely good results on hand-segmented data, which deteriorate by about 10% on automatically segmented images. Including meta-knowledge about vessel segments between two crossings we are able to classify 95.32% of the vessel center lines correctly by combining the ROI feature vector composition method with a neural network. Errors to the greatest part result from the low quality of segments in the outer parts of the image and to very short vessel segments that do not contain enough information for a reliable classification due to the limited

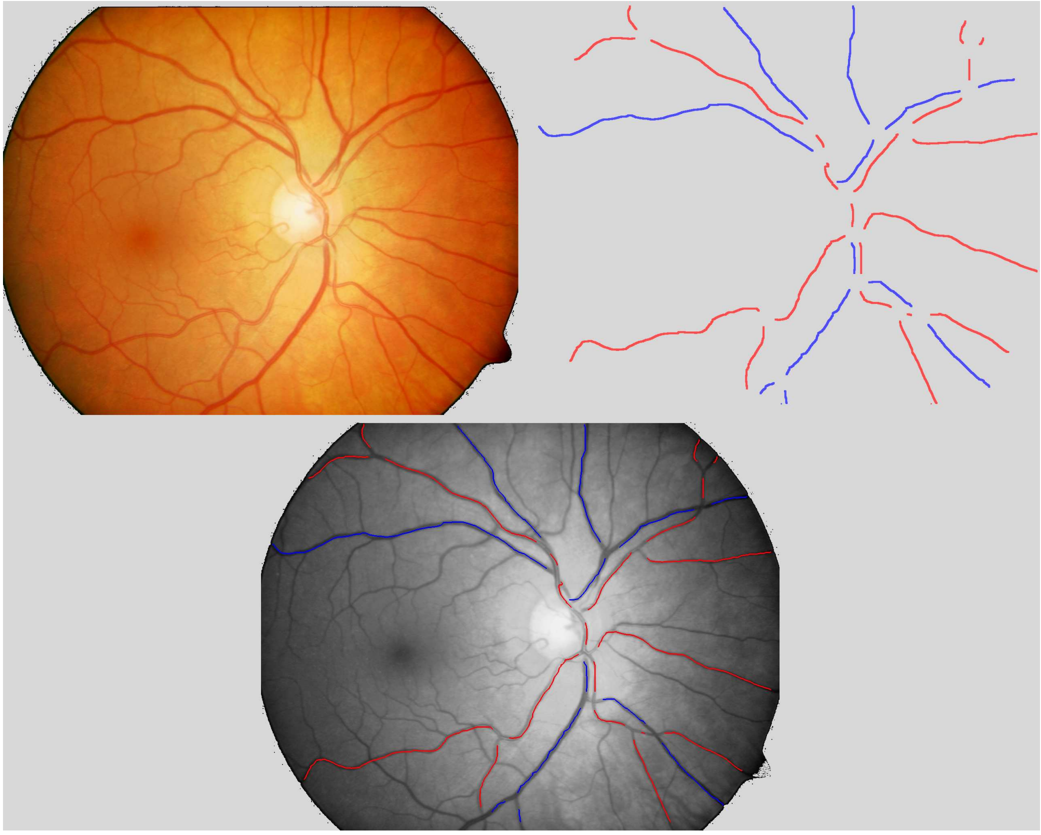


**Figure 9.** Classification result, upper left: original image, upper right: classified vessels (red: correctly classified arteries, blue: correctly classified veins, yellow: vein classified as artery), bottom: superposition of classification result and original green channel

number of samples from this section. Therefore a better vessel tracking algorithm could avoid all remaining misclassifications. Finally, we assume that by applying our approach the classification issue can be understood as solved as soon as segmentation and vessel tracing yield good results.

## REFERENCES


1. J. Leandro, R. Cesar, and H. Jelinek, "Blood vessels segmentation in retina: Preliminary assessment of the mathematical morphology and of the wavelet transform techniques," *XIV Brazilian Symposium on Computer Graphics and Image Processing* , p. 84, 2001.
2. L. Gang, O. Chutatape, and S. Krishnan, "Detection and measurement of retinal vessels in fundus images using amplitude modified second-order gaussian filter," *IEEE Transactions on Biomedical Engineering* **49**, pp. 168–172, 2002.
3. A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Transactions on Medical Imaging* **19**, pp. 203–210, 2000.
4. H. Li, W. Hsu, M. Lee, and H. Wang, "A piecewise gaussian model for profiling and differentiating retinal vessels," *IEEE International Conference on Image Processing* , 2003.
5. E. Grisan and A. Ruggeri, "A divide et impera strategy for automatic classification of retinal vessels into arteries and veins," *International Conference of the IEEE EMBS* , 2003.
6. C. Nieuwenhuis and M. Yan, "Knowledge based image enhancement using neural networks," *18th International Conference on Pattern Recognition (ICPR)* , pp. 814–817, 2006.