Automatic Detection of Anatomical Regions in Three-Dimensional Medical Images

A dissertation submitted to
the Budapest University of Technology and Economics
in fulfilment of the requirements
for the degree of Doctor of Philosophy (Ph.D.)

by

Márton József Tóth

Department of Control Engineering and Information Technology
Budapest University of Technology and Economics

Supervisor:
Balázs Csébfalvi, Ph.D.

Industrial Supervisor:
László Ruskó, Ph.D.

2020
Budapest
## Contents

1. Introduction .................................................................................................................. 1

2. The Data Sets .................................................................................................................. 3

3. Related Work .................................................................................................................. 7
   3.1 Conventional Approaches ......................................................................................... 7
   3.2 Deep Learning Approaches ...................................................................................... 9
   3.3 Detailed Description of the Most Related Approaches ............................................. 9
      3.3.1 Overview of the Classical Approach .................................................................. 9
      3.3.2 Overview of the Neural Network Approach ...................................................... 10
   3.4 Earlier Attempts ......................................................................................................... 12
   3.5 Related Work Summary ............................................................................................ 14

4. Method Overview ............................................................................................................ 15

5. Initial Classification ........................................................................................................ 19
   5.1 Method – First Version ............................................................................................ 19
      5.1.1 Image Pre-Processing ......................................................................................... 19
      5.1.2 Feature Extraction ............................................................................................. 21
      5.1.3 The Zernike Transform ...................................................................................... 21
      5.1.4 Principal Component Analysis .......................................................................... 24
      5.1.5 Machine Learning ............................................................................................. 25
      5.1.6 Training the Classifier ....................................................................................... 28
      5.1.7 Accuracy of the Initial Slice Classification ........................................................ 30
   5.2 Method – Second Version .......................................................................................... 33
      5.2.1 Image Pre-Processing ......................................................................................... 33
      5.2.2 Neural Networks ............................................................................................... 34
      5.2.3 Convolutional Neural Networks ....................................................................... 43
      5.2.4 Initial Classification ........................................................................................... 46
1 Introduction

In modern clinical practice three-dimensional medical imaging has an important role. Different imaging modalities like Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) are routinely applied in diagnostics, radiotherapy planning, tumor monitoring and in many other clinical workflows. By the analysis and evaluation of the collected data, radiologists and other therapists can suggest special treatments that can support the recovery of patients.

In recent years, the number of the acquired images has been increasing rapidly [1]. More and more images are captured and even more resources are used to evaluate these images. To support the work of the radiologists several automatic image processing methods have been developed. These automatic tools are used in various clinical workflows and are usually specialized to a specific region of interest (ROI). There are different applications for lung nodule detection, virtual colonoscopy, cardiac or cerebral vessel analysis or even a simple visualization tool can have different settings for different anatomical regions.

The success of these automatic tools depends on the correct identification of the ROI they can work on. The DICOM standard involves tags to specify the anatomy location of a medical image, but the corresponding field is rarely filled in by the scanner and the manual specification is time consuming and error prone. According to Güld [2] 15% of the images are affected by an incomplete or even wrong indication of the represented anatomy location.

It is easy to see that an automated pre-processing function cannot rely on the DICOM tag only as it may lead to an incomplete processing of the medical image. Such errors can lead to a time-consuming error-correction process that requires serious efforts from the therapist.

It would be advantageous to define a method that can reliably recognize the anatomical regions in medical images based only on the image content itself. It could prevent several errors in the automatic pre-processing steps and therefore would save significant time for the therapists. Beside of this benefit, it would also provide several potential automation or optimization possibilities in numerous clinical applications.
The usage of such a method would be transparent for the therapists and therefore the need for it does not directly comes from the users, but its benefits would be definitely welcome in medical image processing.

In this thesis work, a method is presented to classify the axial slices of three-dimensional medical images according to the anatomical region they belong to. The proposed method has two variations. The first one uses traditional image processing, feature extraction and machine learning tools to provide an initial labeling. The second version replaces these steps and uses a convolutional neural network for this task. As the initial labeling, in both cases, is unreliable and contains major classification errors, a post-processing is applied to ensure a correct and continuous labeling.

This research project started in 2010 when neural networks and deep learning had not yet received much attention in computer vision or in other research communities. The well-known paper of Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton the “ImageNet Classification with deep convolutional neural networks” [3] was two years ahead and today’s deep-learning tools and frameworks were even farther from availability. That time manual image normalization, feature extraction and non-deep machine-learning tools were used to solve image recognition tasks. No well suited workflows were defined and the correct choice and fine tuning of these components required serious effort and research. As the project made progress, I came across deep learning and recognized its importance in the solution of this task. However I did not abandon the original way of the solution. Instead I decided to integrate deep learning in the workflow and created an alternative way to obtain the required classification. I applied the solutions to different imaging modalities and examined their usability and performance on multiple data sets.
2 The Data Sets

In this thesis work, two data sets were used to evaluate the proposed methods. By date the first MRI data set was obtained earlier. It contains fewer cases and also the number of the defined regions are fewer. It was used to develop the first version of the method and later when it became clear that the proposed workflow could handle the recognition task reliably, a more detailed CT data set was assembled. In this way the method could be challenged with more regions to recognize and, as the data were collected in clinical practice, a more detailed overview could be obtained about its usability in real-world cases.

The first data set contained 49 full body MRI cases. It was acquired for research purposes by GE Healthcare in the University Hospital Zurich. The acquisitions were performed with a Discovery MR750w MRI scanner using the same protocol (LAVA-Flex sequence, T1 weighted, FA/TR = 5°/3.7 ms, acquisition matrix = 256x128, 75% phase FOV, scan time 17 s, TE1/TE2 = 1.15 ms/23 ms). The images included all major anatomical regions, the head and neck, the chest, the abdomen, the pelvis, and some part of the legs as well. The test images involved male and female patients of different age (adults only) and level of obesity. The axial resolution of each image was 512x512 pixels. The number of slices varied between 260 and 864 slices (average 447). The pixel size was between 0.39 mm and 1.37 mm (average 0.91 mm), the slice thickness was between 0.47 mm and 8.8 mm (average 5.79 mm). The applied T1 protocol introduces some modality specific image properties, but it does not reduce the usability of the method or the data set as T1 or some similar protocol is available for all MRI devices and its acquisition is part of the regular MRI examination process.

This MRI data set was manually labeled. The axial slices were annotated as LEG, PELVIS, ABDOMEN, CHEST or HEAD&NECK. The definition of these classes was based on the basic structure of the human body. The LEG begins at the bottom of the MRI scan and ends where the two legs join, which is the beginning of the PELVIS. The ABDOMEN begins at the top of the pelvic bone and ends at the top of the liver that is the beginning of the CHEST. The HEAD begins at the top of the shoulders and ends at the top of the head. The test dataset involved 21401 manually labeled axial slices. Based on the manual
labeling the average size and the variance were calculated for each region, which is presented in Table 1.

<table>
<thead>
<tr>
<th>Region</th>
<th>Average Size (mm)</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEG</td>
<td>179.56</td>
<td>85.31</td>
</tr>
<tr>
<td>PELVIS</td>
<td>206.73</td>
<td>27.19</td>
</tr>
<tr>
<td>ABDOMEN</td>
<td>207.83</td>
<td>38.47</td>
</tr>
<tr>
<td>CHEST</td>
<td>188.39</td>
<td>33.82</td>
</tr>
<tr>
<td>HEAD&amp;NECK</td>
<td>234.08</td>
<td>16.78</td>
</tr>
</tbody>
</table>

1. Table. Size of the anatomical regions in the MRI data set.

Table 2 presents statistics about age, height, and weight of the scanned patients. The data set contained 20 female and 29 male patients.

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>Max</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (Years)</td>
<td>18</td>
<td>91</td>
<td>60</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>150</td>
<td>194</td>
<td>169</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>43</td>
<td>127</td>
<td>74</td>
</tr>
</tbody>
</table>

2. Table. Patients' statistics in the MRI data set.

The second data set contains CT examinations. They were originally collected to evaluate various types of clinical applications in the area of oncology, cardiology, surgery, neurology. The cases involve whole-body as well as partial scans. The images were acquired with whole (50-70 cm) as well as small (20-35 cm) reconstruction diameter. The dataset shows significant variance in patient's sex, age, and level of obesity. It includes native and contrast-enhanced cases, and exams demonstrating various types of abnormality (pathology, implant, noise). The number of slices varies between 261 and 874 with 363 average. The slice thickness is between 1.25 mm and 3.75 mm (average 3.12 mm). Thus, it is a representative set of 480 CT images with 92878 axial slices in total.
As, it will be shown later in this work, the recognition of axial slices in the MRI dataset worked reliably, therefore the CT dataset was annotated with more labels. The axial slices were divided into eleven different regions. The definition of the anatomy regions were driven by the needs of organ segmentation algorithms (e.g. brain, lung, liver, colon, prostate). In agreement with radiologists the regions were defined as shown in Table 3. The regions are enumerated in their anatomical order.

<table>
<thead>
<tr>
<th>Region</th>
<th>Starting Point</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOOT</td>
<td>bottom of foot</td>
</tr>
<tr>
<td>SHIN</td>
<td>articular cavity of the ankle</td>
</tr>
<tr>
<td>THIGH</td>
<td>articular cavity of the knee</td>
</tr>
<tr>
<td>PELVIS-LOWER</td>
<td>separation of legs</td>
</tr>
<tr>
<td>PELVIS-CENTER</td>
<td>bottom of symphysis</td>
</tr>
<tr>
<td>PELVIS-UPPER</td>
<td>articular cavity of the hip</td>
</tr>
<tr>
<td>ABDOMEN-LOWER</td>
<td>top of the iliac crist</td>
</tr>
<tr>
<td>ABDOMEN-UPPER</td>
<td>bottom of the lateral sinuses</td>
</tr>
<tr>
<td>CHEST</td>
<td>top of the liver</td>
</tr>
<tr>
<td>HEAD</td>
<td>cranial end plate of the first thoracic vertebra</td>
</tr>
<tr>
<td>BRAIN</td>
<td>foramen magnum</td>
</tr>
</tbody>
</table>

3. Table. Region definitions.

Each region starts at the axial slice involving the starting point and ends before the first slice of the next region. The above defined anatomy landmarks were manually defined by radiologists for all examinations. One exam was processed by one expert, therefore no cross-checking was available. Each slice in the dataset was automatically labeled with the corresponding region label based on its location compared to the expert located landmark points.
Table 4 shows statistics about the regions defined in the CT data set. Different regions have different average length as the corresponding anatomies varies in the human body. The largest differences between the average and minimum values can be seen in case of the FOOT and the SHIN regions. This is caused by the clinical practice. The legs of the patients are usually not scanned in whole unless they are examined directly. In other cases only a smaller part of them is scanned near the examined region.

<table>
<thead>
<tr>
<th>Name</th>
<th>Average (mm)</th>
<th>Max (mm)</th>
<th>Min (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOOT</td>
<td>125</td>
<td>200</td>
<td>23</td>
</tr>
<tr>
<td>SHIN</td>
<td>292</td>
<td>284</td>
<td>35</td>
</tr>
<tr>
<td>THIGH</td>
<td>260</td>
<td>260</td>
<td>221</td>
</tr>
<tr>
<td>PELVIS-LOWER</td>
<td>55</td>
<td>80</td>
<td>13</td>
</tr>
<tr>
<td>PELVIS-CENTER</td>
<td>45</td>
<td>64</td>
<td>30</td>
</tr>
<tr>
<td>PELVIS-UPPER</td>
<td>141</td>
<td>162</td>
<td>101</td>
</tr>
<tr>
<td>ABDOMEN-LOWER</td>
<td>140</td>
<td>210</td>
<td>97</td>
</tr>
<tr>
<td>ABDOMEN-UPPER</td>
<td>64</td>
<td>125</td>
<td>40</td>
</tr>
<tr>
<td>CHEST</td>
<td>175</td>
<td>252</td>
<td>105</td>
</tr>
<tr>
<td>HEAD</td>
<td>119</td>
<td>141</td>
<td>92</td>
</tr>
<tr>
<td>BRAIN</td>
<td>91</td>
<td>160</td>
<td>63</td>
</tr>
</tbody>
</table>

4. Table. CT database region size statistics.
3 Related Work

3.1 Conventional Approaches

As early as 1998 the necessity of a well-defined anatomical knowledge representation has been recognized [4]. This idea has revealed that an anatomical ontology, that defines the physical objects and spaces that constitute the human body, specified in a machine-readable way, could serve well in an automatic medical image-processing workflow as a valuable source of anatomical information.

In the current medical Picture Archiving and Communications Systems (PACS) large amount of images are stored. In these systems, the automatic extraction of anatomical information could increase the efficiency of the automatic image indexing and retrieval operations, which is an essential step in image reuse. According to Florea et al. [5], the reuse of medical images in a context other than a single patient, is infrequent, although several cases with similar medical records are available and even teaching files with images are present in these system. Florea et al. have compared the automatic image categorization capabilities of different PACSs and have found that these systems are capable to recognize the main anatomical structures even in a complex environment. However, the reliability of the recognition has been highly various among the different body regions. This study suggests that the reusability of the archived images will increase if a reliable cross patient data retrieval system is defined, which can further support therapy planning and medical decision making.

The aim of automatic image indexing/retrieval is to provide efficient and fast access to image collections to reuse stored information. When indexing medical images, the automatic categorization provides the means of extracting otherwise unavailable information about the images. Image categorization is usually applied in a context, where a large number of images needs to be treated automatically and where no or only little text is available.

Some research groups have mainly focused on organ or anatomical landmark point detection [6]–[8]. They have used probabilistic algorithms to automatically define salient landmark points that can be used to navigate through the body scans. Their methods have focused on CT scans only, mainly using full body scans.
Other techniques have been developed to estimate the location of the axial slices in the human body scans [9]–[11]. They have used various methods and their accuracies varied between 16.6 mm and 28.3 mm, but they have considered CT modality only.

Atlas based registration methods are also available for anatomy detection. The most related approach [12] performs a non-rigid registration to align a statistical model to a full-body MRI scan, but its usability is limited to full-body MRI scans.

The recent VISCERAL challenge [1] has also showed that the automatic anatomy detection is still a very active field of medical image-processing research and could be well utilized in clinical practice. He et al. [13] have used a region of interest detection for their multi-organ segmentation framework which is a similar task to the anatomy detection. Their work has considered CT modality.

Other works have considered the MRI modality as well. Chen et al. have used their manifold learning method [14] to classify the axial slices of a low-resolution preliminary scan to estimate the patients position. They have achieved a classification result above 90%. A similar approach have been used for slice classification [15] and a 94% overall correct classification rate has been reported.

Criminisi et al. have extended their method [6] to analyze MRI images as well [16] [17]. They have achieved high precision in landmark detection in MRI exams.

To summarize, the existing conventional methods consider CT and MRI images and focus on landmark point or organ detection. A smaller part of the works has focused on the direct classification of the axial slices of a 3D image. It has not received much attention yet and none of the previous algorithms did consider working with partial body images, where the spatially coherent labelling of the axial slices is even more challenging.
3.2 Deep Learning Approaches

Over the last few years, deep learning methods became popular in the computer vision community. The convolutional neural networks (CNN) are especially designed to work with image processing tasks (object detection, classification or segmentation), as images contain highly correlated local intensity patterns. The CNN methods [3], [18], [19] regularly outperform the traditional image feature extraction based methods in image classification tasks.

CNNs have been also applied for recognition of anatomical regions. Yan et al. [20] provided a great overview about the application of different CNN configurations in the classification of axial images. Roth et al. [21] used five classes to identify axial slices of CT images and achieved the state-of-the-art accuracy (5.9% error).

CNNs can be certainly used in other areas of medical image processing. Roth et al. [22] used convolutional networks to segment the pancreas in abdominal CT scans and Liu et al. [23] used CNNs for liver and lung segmentation.

These approaches are just small drops in the ocean of the recent publications using CNNs in many image-processing tasks.

3.3 Detailed Description of the Most Related Approaches

From the above mentioned works there are two that are closely related to this thesis work. The first one has the title “Manifold Learning for Patient Position Detection in MRI” [15] which belongs to the classical approaches and the second one is the “Anatomic-Specific Classification of Medical Images Using Deep Convolutional Nets” [21] that uses convolutional neural network to perform the classification. The second paper was published in 2015.

3.3.1 Overview of the Classical Approach

In [15] the authors have used a special non-linear dimension reduction technique, the Laplacian eigenmaps [24], to map the high dimensional image space into a lower dimensional manifold. They have also considered neighboring axial slices to obtain more accurate results, however they have not performed a whole scan optimization. The classification task itself is performed by a simple k-nearest neighbor method.
They have distinguished six different body regions in their work, namely the HEAD, NECK, LUNG, ABDOMEN, UPPER LEG and LOWER LEG regions. The data set they have used contained 13 whole body MRI scans. Their results are presented in Table 5. They did not publish the whole confusion matrix only the percentage values. The achieved overall accuracy of their classification is 94.0%.

<table>
<thead>
<tr>
<th>Recognized Class Labels in %</th>
<th>true HEAD</th>
<th>true NECK</th>
<th>true LUNG</th>
<th>true ABD.</th>
<th>true U. LEG</th>
<th>true L. LEG</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEAD</td>
<td>95.0</td>
<td>25.9</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>NECK</td>
<td>5.0</td>
<td>69.0</td>
<td>0.9</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>LUNG</td>
<td>0.0</td>
<td>5.2</td>
<td>92.0</td>
<td>2.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>ABD.</td>
<td>0.0</td>
<td>0.0</td>
<td>6.9</td>
<td>97.2</td>
<td>2.1</td>
<td>0.0</td>
</tr>
<tr>
<td>U. LEG</td>
<td>0.0</td>
<td>0.0</td>
<td>0.3</td>
<td>0.8</td>
<td>84.5</td>
<td>1.6</td>
</tr>
<tr>
<td>L. LEG</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>13.5</td>
<td>98.4</td>
</tr>
</tbody>
</table>

5. Table. Results of the manifold learning approach.

Unfortunately, the authors did not publish images with the classification results, similarly to Figure 5, so the region confusion near the region borders cannot be evaluated. However, the achieved overall accuracy is impressive.

In summary, the authors have followed the traditional image recognition approach. They have applied a dimension reduction method, a feature extractor, and applied a classification method. Their results were not filtered by any post-processing method, therefore any possible misclassification would impact the performance of any other methods that uses this information. The data set, they have used, is not a large one, it could have been extended.

### 3.3.2 Overview of the Neural Network Approach

In [21] the authors have used convolutional neural networks to obtain image features and to classify the axial slices of CT images. The network has used five cascaded layers of convolutional filters and two fully connected layers.

The data set contained 4298 selected axials slices, which were retrieved from the PACS of the National Institute of Health. The images were grouped into five regions: LEGS,
PELVIS, LIVER, LUNG and NECK. To extend the available amount of data and to obtain better recognition, the authors have applied data augmentation. They have used random translations, rotations and non-rigid deformation operations on each image to enrich the data set.

The network was trained by a traditional supervised learning process using 80% of the total data set. The rest of the data were used for evaluation. The authors have published the confusion matrix of their results. It is presented in Table 6.

<table>
<thead>
<tr>
<th></th>
<th>true LEGS</th>
<th>true PELVIS</th>
<th>true LIVER</th>
<th>true LUNG</th>
<th>true NECK</th>
</tr>
</thead>
<tbody>
<tr>
<td>pred. LEGS</td>
<td>90</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>pred. PELVIS</td>
<td>0</td>
<td>27</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>pred. LIVER</td>
<td>0</td>
<td>0</td>
<td>518</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>pred. LUNG</td>
<td>0</td>
<td>0</td>
<td>38</td>
<td>88</td>
<td>0</td>
</tr>
<tr>
<td>pred. NECK</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>102</td>
</tr>
</tbody>
</table>

6. Table. Results of the anatomy classification using convolutional networks.

As it can be seen from the table, their overall classification error is 5.9% which is also remarkable and very close to the results mentioned in the previous section. The authors have also published a figure presenting the classification results of a whole body exam. Their method can reliably recognize the NECK, LUNG and LEG regions but the confidence of the neural network greatly drops in the LIVER and LUNG regions. The method does not include any post-processing step.

In summary, this work replaced the traditional way of feature extraction and recognition with a neural network based approach. They have achieved good results but did not extend their work with any post recognition filtering. Their data set was quite limited as well even considering the applied data augmentation process.
3.4 Earlier Attempts

This section does not strictly adhere to present the most related papers of other authors. In these few paragraphs, I will review my previous attempts to solve the slice classification task, in order to illustrate the challenges I faced.

I would like to present here two preliminary approaches that I tried to apply to the problem. The first one used an image similarity metric to compare the axial slices to some reference slices. The labeling was obtained by choosing the most similar reference and assigning the same label to the examined slice. The second approach used the Zernike transform to get an image descriptor. This descriptor was classified by a machine learning tool. To filter this initial classification the whole image series was dynamically divided into regions and in each region a majority vote was generated to obtain the final labeling.

The first approach used the Mutual Information (MI) [25] to calculate the similarity between two images. The idea was to select some reference slices from the available data set that can represent the anatomical regions. The labeling happened by comparing the axial slices of the scans to these reference images and by selecting the most similar slice the resulting label could be used. The similarity was always measured by the MI method.

The first issue that had to be solved in this approach was the selection of the reference slices. To select the references I applied the following method. First, a slice was chosen from an image series. Then the other image series were examined to find the most similar slice in each series. Finally, a score was assigned to the selected slice by counting how many most similar slices were labeled to belong to the same region as the selected one. By applying this method to all slices, a score table could be obtained and from each region the slice with the highest score could be selected as the reference one.

This method was evaluated using a subset of the MRI database. Unfortunately, the method was not able to detect the regions reliably. In some cases the classification was accurate, but in other cases the regions were completely missed.

This could happen because the MI was not be able to capture semantic image content. Only image structures were considered and the variability in the human patients and
the differences in the imaging processes made the recognition process unstable. The MI on its own was not able to solve the problem.

The second approach was published in [26]. In the first step, an image descriptor feature was generated for all slices in the image series using the Zernike transform (for details about the transform see Section 5.1.3). Then a machine learning tool (Bayesian network [27]) was trained to recognize the slices and to assign initial labeling. The result obtained this way was similar to the one presented in Section 5.1.7, but the overall accuracy was only 89.2%. After the initial labeling was obtained for all slices, a post processing step was applied on the whole image series, based on the idea of the previous approach that used MI as a similarity metric.

In an iteratively refining process the image series was divided into five regions. Initially the series was evenly divided. In each region, the center slice was calculated by using the MI. To perform this calculation, in each region, each slice was compared to the other ones in that region. For each slice the sum of the MI calculations was stored. Then the slice with the highest sum was selected as the region center. In the next step of the process the region borders were relocated. A region border was located where, considering neighboring slices, one is more similar to the previous region center and the other slice is more similar to the next center. After the border calculations the iteration could be restarted using the new region division. This process continued while the region borders were changed during iteration.

When this process stopped, a majority vote was applied for each region using the initial labeling. Each region got the label that the majority of the participating slices had.

By this process the original 89.2% accuracy was increased to 90.18%. This is not a major improvement, but it eliminated most of the false recognitions and some alternating initial labeling near the region borders. However, this method had its limitations. It had troubles to label partial body scans accurately and sometimes had serious region border misplacements considering the gold truth labeling.

I think these problems originated from the selection of the not appropriate machine learning tool and from the post-processing method that did not consider the initial
labeling but interpreted the image content in another way than the machine learning tool did.

I think that the above mentioned methods clearly illustrates the difficulties that this task had. Several image descriptor, image similarity metrics and machine learning tools were available to use, but no or only little information was available which one to choose or how to combine them in a system to get a reliable solution.

3.5 Related Work Summary

In summary, the classification of the axial slices of CT and MRI images is widely being researched. The recent CNN classification methods provide state-of-the-art quality results. However, the CNNs use statistical models and can produce noisy output, which cannot be directly used in the clinical practice. Therefore it is worth incorporating a priori information of the human body to stabilize the results of the CNN and to provide a reliable classification.
4 Method Overview

The presented work offers a method to assign labels to the axial slices of three-dimensional medical images according to the body part they belong to. The method has three main parts. The first one is an image pre-processing and normalization step, the second part provides an initial classification while the third step filters and finalizes the results.

The workflow starts with a classical image-processing step. In this phase the axial slices are normalized and the most interfering imaging artifacts are removed. As two different data sets are used in this work, this pre-processing step is modality dependent. Different normalization functions are used for the MR and CT images.

To calculate the initial classifications, two solutions are presented. The first one uses the tools of the traditional image processing methods while the second one relies on modern machine learning and uses convolutional neural networks. The output of both approaches have the same format but have different artifacts and accuracies.

The third step is independent from the underlying workflow elements. It takes the initial classification and the provided classification confidence values and filters the outliers and guarantees a continuous and anatomically correct labeling by fitting region membership functions.

An overview of the proposed method is presented in Figure 1. As it was stated, the method can be divided into three main parts. In step 1, the axial slices of the image are separately pre-processed. This is required to eliminate several disturbing image artifacts and to reduce image complexity. A normalization is also introduced, so the slices can be used in a uniform way in the further processing steps.

Steps 2-4 represent the image recognition part that uses 2-dimensional feature extraction and a machine-learning approach to classify the individual axial slices of the image by assigning the probability of each label to each slice.

In method 3(a), to calculate the feature vectors, a global shape feature extraction algorithm is applied, called the Zernike transform. It is followed by a feature vector size reduction method using Principal Component Analysis (PCA). This way a compact
representation is generated for each slice. These feature vectors are further processed by a Random Forest classifier. This machine-learning approach assigns a probability for each label to an input slice. The parameters of the PCA transform and the classifier are estimated during the training process.

Method 3(b) realizes the same functionality, but uses a convolutional neural network that integrates all the above mentioned features.

Step 4 applies the previously trained model to the test images.

Finally, in the 5th step, a post-processing method is applied. A three-dimensional coherence inspection method, which takes the valid sequence and the mean size of the anatomy regions into consideration to eliminate the false slice classifications, guarantees the correct order of labels and the valid size of the recognized anatomy regions.
1. Figure. The schematic overview of the method.
The presentation of this classification workflow starts with a classical approach. This version uses the MRI data set and utilizes traditional machine-learning tools to provide the initial classification. Section 5 introduces the applied methods and presents the achieved results considering the initial classification.

Following these parts, the second approach is discussed. This version uses the CT data set with a convolutional neural network to provide initial classification. In section 5.2 the corresponding image normalization part is presented followed by the applied neural network topology and the achieved results.

In section 5.3, the two initial classification method are compared not just in numbers, but taking other development considerations into account as well.

In section 6 a post-processing is presented that can be applied after both initial methods. This process filters the initial classifications to meet some anatomical constraints. The post-processed results are evaluated in section 7.

Section 8 summarizes my conclusions about this work, section 9 contains acknowledgments and section 10 presents my thesis points.
5 Initial Classification

5.1 Method – First Version

This section introduces the initial classification using a conventional image-processing method and the MRI data set. The main idea of this part was published in [28], [29], [30] and [31].

5.1.1 Image Pre-Processing

MR images acquired in the clinical practice are usually retrieved from different scanners and show significant variation in patient size, positioning, and have different image properties (such as intensity range and image resolution). These differences originate from the different imaging settings of the scanners and the anatomical differences of the patients.

The first thing that has to be normalized is the intensity range of the image. Since the MRI intensities are not standardized like in case of CT images, the intensity range can vary on a large scale, even if they were acquired using the same (e.g. T1) protocol. To deal with this phenomenon the intensity range of the MRI scan is normalized to the range $[0, 255]$ using a histogram-based and a linear technique. First, the histogram of the axial slice is computed then the $98\%$ percentile value is computed. This provides an upper limit to the intensity values and filters the high value outliers. After the computation of the upper limit, the recalculated intensity scale is mapped to the normalized range. After this normalization the main tissue types have similar intensities in different MR images, so the image can be handled uniformly in the further processing.

Differences caused by the different imaging settings have to be also handled carefully. The input images may have different resolution, field of view (FOV), and even the positioning of the patients can vary. To handle these problems a multistep process is applied. First, the resolution of the axial slices is normalized. The input slices are down-sampled to $128 \times 128$ pixels. This down-sampling reduces the latter computation time significantly and it still preserves the information required to recognize an axial slice.
The next step of the normalization is the correction of the patient position. Different positions can cause translation of the body on the axial slices. To correct this effect, the Hu image moment [32] is computed for each axial slice, as it is showed by Equation (1):

\[ m_{ipq} = \sum_x \sum_y x^p y^q f_i(x, y). \]  

(1)

In Equation (1) \( f_i(x, y) \) defines the image function belonging to the \( i \)-th axial slice. To correct the translation effect, the center of gravity of each axial slice has to be moved to the center of the image. This can be done by applying a transform from \( f_i(x, y) \) to \( f_i(x + (x^c - x^s), y + (y^c - y^s)) \), where, \( x^s = m_{i10}/m_{i00} \), \( y^s = m_{i01}/m_{i00} \) and \( x^c \) and \( y^c \) denote the coordinates of the center of the image.

After the translation correction, a FOV normalization is applied. In this step, the image is cropped to a fixed physical size. This ensures that the cross section of the body fills the image disregarding the original pixel size and image resolution (that is always available in the DICOM header of the input images). The result of this step together with the other main steps of the pre-process is shown in Figure 2.

![Figure 2](image.png)

2. Figure. Slice normalization: the original slice (a); the result of the translation correction (b); FOV normalization (c).

After this normalization process the input image is free from the most interfering artifacts and can be used in the further steps of the proposed method.
5.1.2 Feature Extraction

As the resolution of the original axial slice can be 128x128 pixels, it can be considered as a more than 16 000 dimensional vector. It means that the primary classifier has to find an optimal subdivision of a 16 000 dimensional space to represent the five classes of the anatomical regions. This task would require a complex structure with many trainable parameters. To reduce this complexity, the idea is to reduce the dimension of the input while maintaining its information content. This procedure is called feature vector extraction.

In this version of the method, to represent the structural content of an axial slice and to make it suitable for classification by a machine-learning technique, a reduced number of features is extracted. The computation of these feature vectors consists of two steps. First, a global image shape descriptor is applied to capture the structural content of the image and subsequently a dimension reduction method is applied to the generated feature descriptor to minimize its size and the complexity of the machine learning model. In the first step, the Zernike transform is used, while in the second step a Principal Component Analysis (PCA) is applied.

5.1.3 The Zernike Transform

The original mathematical transform has been introduced by F. Zernike [33] for optical problems and his work has been later applied in the field of image processing [34] for image feature extraction. The Zernike transform provides a set of orthogonal complex moments of an image. These moments can be used as a global image descriptor. The moments can capture the structural content of the image, so they are widely used as global scene/shape descriptor [35][36][37] when the displayed shape is a fundamental property of an image. There are several shape descriptors that are used in image retrieval and classification tasks, but the Zernike transform has some important properties which can be exploited in the recognition of the axial slices of a 3D medical image. The transform can be formalized in a rotation invariant way, which is an important property as it makes the transform not sensitive to the rotation of the patient. The Zernike transform performs better in shape recognition tasks than other geometrical moments [38] and it shows resistance to some typical characteristics of MRI scans like the unstandardized intensity or the intensity inhomogeneity.
According to its mathematical formulation the transform is interpreted within a unit disk and this geometrical arrangement is similar to the cross-section of the human body and the circular shape of the FOV of the axial slices in 3D medical images.

The Zernike moments can be expressed as a set of complex polynomials $V_{mn}(x, y)$ over the interior of the unit disk i.e., $x^2 + y^2 \leq 1$. In polar coordinates the form of the polynomials is:

$$V_{mn}(x, y) = V_{mn}(r, \theta) = R_{mn}(r) \exp(jm\theta),$$

where $n$ is a non-negative integer, $m$ is an integer, $n - |m|$ is even and $|m| \leq n$; $r = \sqrt{(x - cx)^2 + (y - cy)^2}$ is the length of the vector from the center of the image $(cx, cy)$ to a pixel $(x, y)$; $j$ is the complex unit vector and $\theta = \arctan(y/x)$ is the angle between the vector $r$ and the $x$ axis. $R_{mn}$ is the Radial polynomial defined as:

$$R_{mn}(r) = \sum_{k=0}^{(n-|m|)/2} \frac{(-1)^k(n - k)!}{k! ((n + m)/2 - k)! ((n - m)/2 - k)!} r^{n-2k}.$$  \hspace{1cm} (3)

The Zernike moments are the projection of the image function onto these basis functions. So the two-dimensional Zernike moments of order $n$ and repetition $m$ for a discretized function $f(r, \theta)$, which vanishes outside the unit disk, can be defined as follows:

$$Z_{mn} = \frac{n + 1}{\pi} \sum_r \sum_\theta f(r, \theta) V^*_m(r, \theta),$$

where $V^*_m(r, \theta) = V_{n-m}(r, \theta)$. In this work, the $f(r, \theta)$ function represents the image (i.e. a 2D axial slice).

Suppose that one knows all the moments $Z_{mn}$ of the function $f(r, \theta)$ up to a given order $N_{\text{max}}$, to reconstruct a function $\hat{f}(r, \theta)$ whose moments exactly match those of $f(r, \theta)$ up to the order $N_{\text{max}}$ one should use the following equation:

$$\hat{f}(r, \theta) = \sum_{n=0}^{N_{\text{max}}} \sum_{m} Z_{mn} V_{mn}(r, \theta),$$

where the same constraints are applied as in (2). Note that as $N_{\text{max}}$ approaches infinity, $\hat{f}(r, \theta)$ will approach $f(r, \theta)$. 

22
Figure 3 illustrates the results of the forward transform followed by a back transform using different orders of the Zernike transform. In Figure 3(a) one can see the original axial slice, while 3(b), 3(c), and 3(d) show the reconstructed images using 5th, 25th, and 50th order transform, respectively. Applying a higher order transform captures more details but increases the number of the moments rapidly, therefore the order of the transform shall be limited. Based on visual inspection of MRI images the 25th order of the transform was selected, which provides a visually recognizable reconstructed image while the number of the provided moments is still manageable.

(a) Original Image.  
(b) Reconstructed from the 5th order transform.  
(c) Reconstructed from the 25th order transform.  
(d) Reconstructed from the 50th order transform.

3. Figure. Reconstruction from different orders of the Zernike transform.
5.1.4 Principal Component Analysis

PCA [39] is a mathematical transform that uses an orthogonal transform to map a series of observations of correlated variables into an uncorrelated set of variables. These variables are called principal components. PCA is used for dimension reduction while the variance of the samples is preserved as much as possible. The procedure is defined in a way that the first principal component has the highest variance and each succeeding component has the largest possible variance under the constraint that it has to be orthogonal to the previous components.

To calculate the principal components the following method has to be performed. In the first step, the input data has to be normalized to avoid bias to some variables:

$$F_{ni} = \frac{F_i - M_i}{sd_i},$$  \hspace{1cm} (6)

where $F_i$ is the i-th component of a multi-dimensional feature vector, $M_i$ is the mean value of that component considering the whole data set and $sd_i$ is the standard deviation of the i-th component.

The second step of the PCA calculation is the construction of the covariance matrix. This matrix measures the joint variability between of the components of the multi-dimensional $F_n$ feature vector:

$$\begin{bmatrix}
Cov(F_{n1}, F_{n1}) & Cov(F_{n1}, F_{n2}) & \ldots & Cov(F_{n1}, F_{nn}) \\
\vdots & \vdots & \ddots & \vdots \\
Cov(F_{ni}, F_{n1}) & Cov(F_{ni}, F_{n2}) & \ldots & Cov(F_{ni}, F_{nn}) \\
\vdots & \vdots & \ddots & \vdots \\
Cov(F_{nn}, F_{n1}) & Cov(F_{nn}, F_{n2}) & \ldots & Cov(F_{nn}, F_{nn})
\end{bmatrix}.$$  \hspace{1cm} (7)

The covariance between two real valued random variable $X$ and $Y$ is defined as:

$$cov(X, Y) = E[(X - E[X])(Y - E[Y])],$$  \hspace{1cm} (8)

where $E[X]$ and $E[Y]$ are the expected values of the $X$ and $Y$ variables respectively. These theoretical values can be estimated on the data set and the covariance can be calculated for the feature vector components. As $cov(X, Y) = cov(Y, X)$ the covariance matrix is symmetric.
The third step of the method is the calculation of the eigenvalues and the corresponding eigenvectors of the covariance matrix. To perform these calculations different methods can be used (Cholesky decomposition [40] or QZ algorithm [41]).

The resulted eigenvectors act as the principal components we are looking for and the eigenvalues measure the variance of the data set along that principal component. Thus, if the eigenvectors are ordered by their eigenvalues in descending order, the significance of the principal components can be found.

As there are as many components as the number of variables in the data, principal components are constructed in such a manner that the first principal component has the largest possible variance in the data set. By limiting the number of principal components, discarding the last \( k \) components, one can effectively reduce the size and dimensions of the data set while preserving as much information as possible.

Of course this approach introduces some information loss, but this is a balancing between the amount of data and accuracy.

Although, the Zernike transform captures all the necessary information, the 25\(^{th}\) order transform still provides near 200 complex moments. It has to be reduced to get a feature vector that can be handled by a machine learning tool. The transformation matrix of the PCA is calculated on the training data set, as this calculation is part of the training process. Later, this pre-calculated transformation is applied in the recognition phase to reduce size of the output of the Zernike transform approximately to the tenth of its original size. In this work the WEKA data mining tool [42] was used to perform the calculation of the PCA transform.

By applying this two-step feature vector extraction process, a compact representation can be obtained for each axial slice of an MRI scan. Using this representation, a machine learning tool can be trained to assign anatomy label to each axial slice of a 3D image.

5.1.5 Machine Learning

In many image recognition works various machine learning tools are applied to classify images based on extracted features. In this work the Random Forest classifier [43] was used. This classifier creates multiple, independently trained decision trees. It was shown
in [43] that this composition of decision trees has better generalization ability than the classical decision trees through the applied randomness during the training process.

Supervised, discriminative classification methods have been successfully applied in many fields of medical image processing like tumor detection [44] or organ localization [45]. Support Vector Machines (SVM) [46], AdaBoost [47] and Probabilistic Boosting Trees [48] are well known and proved methods for these tasks. However, it has been shown that random forest performs better in multi-class problems than SVMs and it is more effective than boosting [49] [50].

The Random Forest classifier is basically a set of decision trees. To train a conventional binary decision tree, the following approach is used [51].

First assume that there is only one variable $X$ and a binary classifier is built. To create the tree, first a split point $t$ is chosen that divides the input space into two sets $A_1 = (-\infty, t]$ and $A_2 = (t, \infty)$. Let $\bar{Y}_1$ be the mean of $Y_i$ labels in the training data set that are in $A_1$ and $\bar{Y}_2$ be the mean of $Y_i$ labels in the training data set that are in $A_2$.

For a binary classifier the split is chosen to minimize a surrogate for classification error. A common choice is the impurity function: $I(t) = \sum_{s=1}^{2} \gamma_s$ where

$$\gamma_s = 1 - \left( \bar{Y}_s^2 + (1 - \bar{Y}_s)^2 \right). \tag{9}$$

This impurity function is known as the Gini index [52]. If a partition range $A_s$ contains all the training set elements that are 0 or 1, then $\gamma_s = 0$, otherwise $\gamma_s > 0$. In the training process $t$ is chosen to minimize the impurity. Other impurity metrics, such as entropy, are also frequently used in practice. The impurity function is used rather than the classification error because impurity is a smooth function, therefore it is easy to minimize.

At this point the splitting can be continued recursively until a stopping condition is met. Such condition can be a predefined maximum depth of the tree or if the partition ranges $A_s$ contain fewer elements than a limit. The bottom nodes of the tree are called leaves. In each leaf the mean $\bar{Y}$ of $Y_i$ labels can be calculated. For classification if $\bar{Y} \geq \frac{1}{2}$ the leaf classifies the input as 1, 0 otherwise.
For multi-dimensional inputs the split point and variable is chosen to minimize the impurity function during the training process.

Multiclass classification can be solved in the same way, just the impurity function has to be extended as shown in (10):

$$\gamma_s = 1 - \sum_{i=1}^{C} p(i)^2 .$$

In this formula the number of classes is $C$ and $p(i)$ is a ratio showing us, how many of the input points are classified as a member of class $i$. If all inputs are enrolled in the same class by a given node in the decision tree, than the impurity of that node is 0 and some larger positive value otherwise. Based on this metric, an optimal splitting point can be found, that minimizes the impurity of the examined node.

The result of this training process is a piecewise linear constant estimator that can be represented as a tree.

The decision trees are useful and simple classifiers and their prediction capability can be increased by combining many of them. A common approach is called bagging. Suppose we draw $B$ bootstrap samples and each time construct a classifier. This gives $B$ independent decision trees. To classify a single sample each tree preforms the classification task and the class with the most votes is assigned to the sample.

From bagging, the concept of random forests is just a step ahead. Random forests are bagged trees except that only an independently selected random subset of variables is used for each tree training.

By performing bagging and randomization, a complex set of simple classifiers is obtained that can produce more accurate and reliable results than a single decision tree.

The random forest has another important property that can be utilized in this work. It does not only assign a label for each axial slice, but also gives the confidence of its decision. As the random forest consist of multiple decision trees, each tree assigns a label for each axial slice. The distribution of this labeling shows what percentage of the trees voted for each possible label, which can be interpreted as a confidence value for
each label. This confidence distribution of the labels is used in the next step of the proposed algorithm to correct the errors in the result of the initial classification. As all these properties of the random forest were favorable in this work, this technique was selected as the machine-learning tool. An implementation of random forest is also available in the WEKA toolset [42].

5.1.6 Training the Classifier

To calculate the parameters of the PCA and to train the random forest classifier, the following training process was performed (illustrated in Figure 1.). First, the image database is randomly split into two parts. The smaller part (including 22 patients) is used to train the system while the larger part (including 27 images), is used for the evaluation of the trained system. The training process estimates the parameters of the PCA and the random forest and provides a model that is used to label the test data. The initial classification is achieved by assigning the most likely label to each slice. After the labeling is performed, its result is compared with the manually assigned reference labeling. The result of the comparison is displayed using a confusion matrix.

To monitor the progress of the training process and to have a more accurate picture about the overall performance of the classifier a ten-fold cross validation was used during the training process. This means that the training data set was further divided into a concrete training and an evaluation set. The concrete training set contained 90% of the training data. This set was used for the actual training and the performance of the classifier was evaluated on the evaluation set. This partitioning and evaluation was repeated ten times. This way the performance of the classifier could be monitored and their meta parameters could be optimized.
In Table 7 and 8 the training times and the achieved accuracy can be seen using different meta parameter settings for the Random Forest classifier and the PCA transform. One can see that both the number of trees and the number of PCA components had a great impact on the training time and accuracy. Increasing the number of the trees increased both the accuracy and the training time. PCA had different impact. Not applying PCA made the process less accurate and had greater training time, using too much components had the same effect. However, using only 25 components dropped the accuracy in the same way, so a balance had to be found. According to the measurements using 50 PCA components with 500 trees in the random forest maintain good accuracy while keeping the training time in an acceptable range.

After the training process, the trained classifier was evaluated on the original test data set.
5.1.7 Accuracy of the Initial Slice Classification

The confusion matrix of the performance evaluation is presented in Table 9. The columns of the table specify the known ground truth labeling while the rows represent the labeling assigned by the trained system.

To assess the performance of the system, the following metrics were used. For each anatomy label the class precision (last column) was calculated as the number of the correctly labeled slices divided by the number of all slices that are associated with that specific label by the proposed method. The class recall (last row) metric is defined as the number of the correctly labeled slices for a specific label class divided by the total number of slices belonging to that specific class according to the ground truth labeling. The overall accuracy (black cell in last row) is defined as the ratio of the number of correctly labeled slices (belonging to any label) to the total number of slices in all classes.

Table 9 demonstrates the summarized accuracy of the random forest classifier measured on the test set. According to the table the overall accuracy was 91.27%, which means that 91.27% of the slices involved in the test set were correctly labeled. The class recall was the highest (99.06%) for HEAD&NECK, which means that 99.06% of the slices, which belong to the head according to the ground truth, were labeled as HEAD&NECK by the initial classification. The class precision was the highest (97.90%) for HEAD&NECK, as well. It means that 97.90% of the slices, which were classified as HEAD&NECK, did actually belong to the head (according to the ground truth). This means the head region was recognized with the highest confidence. In contrast to HEAD&NECK the PELVIS is recognized with the least confidence (see 82.90% recall and 89.12% precision).

The main diagonal of the table contains the correctly labeled slices for each class. The elements directly above and below of the diagonal represent slices which were confused with a neighboring region, while the remaining elements represent large confusion in the initial labeling. The first type of the confusion errors is not considered as a major problem, because it usually means border shift between the neighboring regions or some alternating labels near the region borders. The second type of confusion is considered as greater mismatch which means the classifier completely missed the labeling (e.g. 50 PELVIS slices were labeled as CHEST by the initial classification). This effect could be caused by several factors. As the random forest classification is a
machine learning tool that performs statistical operations, it has the possibility that the generated model could not handle all feature vector occurrences correctly. Another factor that should be considered is the variation of images including image noise. Figure 4 shows an example where the exam was merged from multiple sub-scans and the original borders contain high level noise, therefore the initial classification fails to recognize those regions. The different colors represent the different anatomy regions (yellow - HEAD&NECK, purple - CHEST, green - ABDOMEN, red - PELVIS, and blue – LEG). The left side of the image contains a coronal view of the exam, showing the merging artifacts, and on the right side the confidence values provided by the Random Forest are plot.

![Figure 4. Image artifacts and initial classification results.](image)

Two other examples for the above-mentioned errors are illustrated in Figure 5. The left side of both examples shows the initial labeling (based on maximum confidence) and the confidence values provided by the random forest are plot on the right side. As it can be seen, even the most accurate classification result (Figure 5(a)) includes some outliers (some slices in the middle of the ABDOMEN are labeled as PELVIS). This indicates that the independent processing of slices, even if the overall accuracy seems high enough, cannot be reliable, as it does not incorporate the neighborhood of the slices. When
Figure 5(b) is considered, the results are more confusing. Although the majority of slices are correctly classified, the result contains several misclassified slices (e.g. chest detected in the pelvis). Some labels are confused with labels of other regions, so such a result cannot be considered as useful information in clinical applications.

The achieved initial classification can be compared with the results of papers [15] and [21]. Their results are presented in Tables 5 and 6 respectively. As it can be seen, the presented approach uses five labels while the others consider six and five regions. The achieved overall accuracy (91.26%) is slightly lower than that the others presented (94.0% and 94.1%). While in [15] some neighboring slices are processed together, the overall behavior is similar in all cases. The independent, or nearly independent, processing of the axial slices makes it hard to obtain a continuous and outlier free labeling.

The results summarized in Table 9 and the observation illustrated in Figure 5 led to the recognition of the necessity of a post-processing step that takes the neighborhood of the slices as well as the normal size of anatomy regions into consideration to assign correct labeling to an image series.

<table>
<thead>
<tr>
<th></th>
<th>true LEG</th>
<th>true PELVIS</th>
<th>true ABDOMEN</th>
<th>true CHEST</th>
<th>true HEAD&amp;NECK</th>
<th>class precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>pred. LEG</td>
<td>1514</td>
<td>81</td>
<td>5</td>
<td>7</td>
<td>0</td>
<td>94.21%</td>
</tr>
<tr>
<td>pred. PELVIS</td>
<td>117</td>
<td>1663</td>
<td>86</td>
<td>0</td>
<td>0</td>
<td>89.12%</td>
</tr>
<tr>
<td>pred. ABDOMEN</td>
<td>0</td>
<td>212</td>
<td>1735</td>
<td>116</td>
<td>1</td>
<td>84.06%</td>
</tr>
<tr>
<td>pred. CHEST</td>
<td>3</td>
<td>50</td>
<td>88</td>
<td>1665</td>
<td>19</td>
<td>91.23%</td>
</tr>
<tr>
<td>pred. HEAD&amp;NECK</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>43</td>
<td>2098</td>
<td>97.90%</td>
</tr>
<tr>
<td>class recall</td>
<td>92.66%</td>
<td>82.90%</td>
<td>90.55%</td>
<td>90.93%</td>
<td>99.06%</td>
<td>91.27%</td>
</tr>
</tbody>
</table>

9. Table. Accuracy of the initial slice classification.
5.2 Method – Second Version

This section introduces an approach that uses a CNN and the CT data set to perform the initial classification. This part of the work was published in [53] and [54].

5.2.1 Image Pre-Processing

The position of the patient with respect to image center can vary significantly among examinations, which introduces unwanted variation. Even if the patient is acquired in supine position, the table can be lifted in anterior or posterior direction. In order to compensate that, the weight center of non-air voxels is computed that is used to define the body axis for the whole CT exam.

After this point, each slice is separately processed. First, a squared region is extracted, such that its center is located on the body center axis (that is defined as the average of the weight center for all slices), its size (both width and height) is equal to 35 cm, and its resolution is equal to 256x256 pixels. Since the pixel size of CT images varies among examinations, interpolation is used. If the original input slice covers larger or smaller image region than 35 cm, the pre-processed image is cropped or padded (with air voxels), respectively. Then, the pixel intensities (from the original 16-bit signed integer
value) are rescaled to the range [0,255], such that the range [-500,500] HU is linearly mapped to [0,255]. This step enhances those tissue types which are key important from anatomy point of view and prevents the classifier taking very high or low density pixels into account.

6. Figure. Pre-process result. (a) Original axial slice, (b) Pre-processed axial slice.

Figure 6 shows the result of the pre-process. As a consequence of this step, the most interfering imaging options can be normalized and the resulted axial slices can be used in a uniform way in the following steps.

5.2.2 Neural Networks

In the recent years, the image classification tasks are usually solved with artificial neural networks, as they offer a more comfortable way to achieve the same or better results compared to the classical image processing approaches.

The first model of the artificial perceptron was introduced by Frank Rosenblatt in 1957 [55]. He described it as a simplified mathematical model of how neurons in the human brain work. They take multiple binary input, multiply each one with a continuous valued weight, sum them and apply a threshold providing 1 output if the weighted sum is above a limit or 0 otherwise. The input of a perceptron can come directly from the input data or it can be the output of other perceptrons. It also has a special input, which always has the value of 1. It is called bias, and it makes it possible to shift the output function. The concept of the perceptron model is illustrated in Figure 7 (the bias is marked with $b$ in the figure).
The weights in the model can be set using a training method. The original idea to find the appropriate weights is pretty straightforward. A subset of the original data set, with input-output pairs, is used in the training. The weights of the perceptron is randomly initialized. Then the output of the perceptron is calculated for a training input element. If the output is 0, but the reference output is 1, than increase the value of the weights until the output is 1, if the output is 1, but the reference is 0, than decrease the value of the weights, otherwise leave the weights unchanged. Then move on to the next element of the training set.

This simple method works well for the perceptron model and it can teach a perceptron to reproduce simple logical functions like AND/OR/NOT. However, a bit more complex XOR function cannot be learned by this model as this problem is not linearly separable (two sets, \(X_0, X_1\), are linearly separable in the \(n\) dimensional space if there exists \(n + 1\) numbers, \(w_1, w_2, \ldots, w_n, k\), such that \(\forall x \in X_0\) and \(\forall x' \in X_1\) \(\sum_{i=1}^{n} w_i x_i > k\) and \(\sum_{i=1}^{n} w_i x_i' < k\) where \(x_i\) and \(x_i'\) are the \(i\)-th component of \(x\) and \(x'\) respectively).

![Perceptron Model](image)

7. Figure. The perceptron model.

To overcome this limitation a more complex architecture is suggested. Individual perceptrons can be grouped into layers and layers can be stacked upon each other. This structure is called Multi Layer Perceptron model, illustrated in Figure 8.

In this structure, instead of having only one output layer, the input can be sent to arbitrary number of neurons that form a hidden layer. A layer is called hidden because its output acts as an input to another hidden layer or to the output layer and it is not directly visible in the output of the network.
The reason why hidden layers are useful in this structure, is that the hidden layers can find features within the data and the following layers can work on these features rather than on the noisy and large original input data.

For example, in an image classification task, the first layer can take the raw input pixels and can extract features like edges, circles, corners etc. The next layer receives the position of these features in the image and can form more complex features from them. This continues through the hidden layers until the last layer can classify the image using very high level features.

![Multi-Layer Perceptron model](image)

8. Figure. Multi-Layer Perceptron model.

At this point we have a multi-layer model and some training data, which specify the input and the expected output at the end of the network. But how should we train the weights inside the network to get the expected result? The training method of Rosenblatt [55] does not work for this model as it can handle only single layers and does not specify what to do with the previous layers.

The solution is provided by the well known derivation and the chain rule. The key realization is that, if the neurons were not perceptrons, but were made to calculate their output with a nonlinear activation function which is also differentiable, derivation can be used to adjust the weights to minimize the error and the chain rule could also be used to compute the derivative for all the neurons in the previous layer and thus their weights would also be known. This method is called backpropagation.

Backpropagation was derived by multiple researchers [56][57] in the 60’s and 70’s but got attention only in 1986 when David Rumelhart, Geoffrey Hinton and Ronald Williams published their paper [58].
To be more formal, the total input, \( x_j \), to neuron \( j \) is a linear function of the outputs, \( y_i \), of the neurons that are connected to \( j \) and of the weights, \( w_{ji} \), on the connections:

\[
x_j = \sum_i y_i w_{ji}.
\] (11)

Neurons can have biases by introducing an extra input to each neuron which always have a value of 1. The weight on this extra input is called the bias and it can be treated as any other weights.

A neuron has a continuous valued output \( y_j \) which is a non-linear function of its total input:

\[
y_j = \frac{1}{1 + e^{-x_j}}.
\] (12)

In this equation the sigmoid function is used, but any non-linear function can be used that has a bounded derivative.

The aim is to find a set of weights that ensures that for each input data the output of the network is close to the expected output value. For a given input-output pair, the error of the network is defined as:

\[
E = \frac{1}{2} \sum_j (y_j - d_j)^2,
\] (13)

where \( j \) is an index over the output neurons, \( y_j \) is the actual output of neuron \( j \) and \( d_j \) is the desired output. To minimize \( E \) by a gradient descent method, it is necessary to compute its partial derivatives with respect to each weight \( w_{ji} \). For a given input-output pair the partial derivatives of the error are calculated in two phases. The first phase is called forward pass in which the output of the neurons in each layer is calculated by (11) and (12). The second phase is called backward pass.

The backward pass starts with the computation of \( \partial E / \partial y \) for each output unit. Differentiation equation (13) gives the following result:

\[
\frac{\partial E}{\partial y_j} = y_j - d_j.
\] (14)

Here the chain rule can be applied to determine \( \partial E / \partial x_j \).
\[
\frac{\partial E}{\partial x_j} = \frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial x_j}.
\] (15)

By differentiating equation (12) and substituting it into (15) gives:

\[
\frac{\partial E}{\partial x_j} = \frac{\partial E}{\partial y_j} y_j \left(1 - y_j\right) = \frac{\partial E}{\partial y_j} \left(e^{-x} \right)^2.
\] (16)

This equation tells, how a change in the total input \(x_j\) of an output neuron affects the error. But \(x_j\) is a linear function of the outputs of lower level neurons, so it is possible to compute how the error will change by changing the weights of the connections or the output of the previous neurons. For a weight \(w_{ji}\), from \(i\) to \(j\) the derivative is:

\[
\frac{\partial E}{\partial w_{ji}} = \frac{\partial E}{\partial x_j} \frac{\partial x_j}{\partial w_{ji}} = \frac{\partial E}{\partial x_j} y_i.
\] (17)

For the output of the \(i\)-th neuron, its contribution to \(\frac{\partial E}{\partial y_i}\) resulting from the connection from \(i\) to \(j\) is:

\[
\frac{\partial E}{\partial x_j} \frac{\partial x_j}{\partial y_i} = \frac{\partial E}{\partial x_j} w_{ji}.
\] (18)

By taking into account all the connection starting from neuron \(i\) \(\frac{\partial E}{\partial y_i}\) is:

\[
\frac{\partial E}{\partial y_i} = \sum_j \frac{\partial E}{\partial x_j} w_{ji}.
\] (19)

At this point, we know how to compute \(\frac{\partial E}{\partial y_i}\) for any neuron in the last but one layer if the \(\frac{\partial E}{\partial y_j}\) is known for all output neurons. This method can be successively applied to the previous layers and neurons and the \(\frac{\partial E}{\partial w}\) can be calculated for each neuron in the network.

By knowing all the \(\frac{\partial E}{\partial w}\) values each weight can be updated iteratively during the training process:

\[
\Delta w = -\frac{\partial E}{\partial w}.
\] (20)

This is the simplest method, when the weights are updated after each input-output pairs. This approach does not require large amount of additional memory, but
computationally more efficient implementations use averaging and perform these operations in batches of input-output pairs.

In this approach, it is important to assemble the batches correctly. The statistical properties of the batches shall be similar to the properties of the whole data set, otherwise the averaging can distort the gradient.

To minimize the effect of the gradient distortion and the possibility to stuck in a local minimum of the cost function, a momentum can be introduced (21). It tries to keep the direction of weight changes intact, it filters small fluctuations and allows only slower changes.

$$\Delta w(t) = -\varepsilon \frac{\partial E}{\partial w(t)} + \alpha \Delta w(t - 1),$$  \hspace{1cm} (21)

where $\alpha$ is a factor between 0 and 1 that describes the contribution of the current and earlier gradient to the actual weight change.

These formulas allows the training of multi-layer networks and the backpropagation method is still an inevitable algorithm in the training process of neural networks.

Soon after the appearance of the backpropagation algorithm, its usefulness was demonstrated in a significant real world problem. At the AT&T Bell Laboratories Yann LeCunn et al. published the paper “Backpropagation Applied to Handwritten Zip Code Recognition” [59]. In this work they have presented a neural network that can recognize grayscale images of single digits.

This was an important working neural model in the filed of image processing. The structure of the proposed network was also interesting. The first hidden layer applied the concept of weight sharing. This means that instead of each neuron having a different weight for each input pixel, only a small set of shared weights is applied on the whole image. Today this type of weight sharing between the neurons is called convolutional layer, which is applied to extract structural features in a computationally more effective way than the fully connected layer would do.
By using the backpropagation method, multi-layer neural networks can be trained. However, the maximum complexity of the trainable networks is limited. Increasing the number of layers leads to a problem called vanishing or exploding gradients.

The backpropagation method calculates the weight changes by applying a series of gradient estimations and multiplications. If we consider a network with multiple consecutive layers, a whole series of multiplication is needed to calculate the weight changes in the first layer. This can cause problems. If the multiplied numbers are below one, than the multiplication series will provide a result close to zero, this is the vanishing gradient problem, or if the values are above one, the result of the multiplications will be a very large number and this causes the exploding gradient problem.

To train deep neural networks these problems have to be handled. To overcome this limitation several considerations have to be taken.

The first approach is the appropriate selection of the activation function. In (12) the sigmoid function is used as nonlinearity. However, if we consider its derivative function, (22) and Figure 9, we can see that it is almost zero if the input is not close to zero. This implies two complications. During the training process the network will converge slowly, if its starting positions is far from the optimum. The second problem is more frustrating, as the low gradient value is the main reason of the vanishing gradient problem, if a sigmoid activation is used in the hidden layers:

$$\frac{\partial y_j}{\partial x_j} = \frac{1}{1 + e^{-x_j}} \left(1 - \frac{1}{1 + e^{-x_j}} \right).$$

(22)

![The sigmoid function.](image.png)

![Derivative of the sigmoid function.](image.png)

9. Figure. The sigmoid function and its derivative.
Instead of the sigmoid function, the rectifier activation [60] can be applied. A neuron using this activation is usually called rectifier linear unit aka ReLu. The rectifier function itself is defined as:

\[ y_j = \max(0, x_j). \]  \hspace{1cm} (23)

This function has constant 1 or 0 derivative if the input is greater than 0 or lower than 0 respectively. Mathematically this function is not differentiable at 0, but at this point its derivative can be replaced with 0 or 1. It does not cause numerical problems. Even if the derivative of this function is 0 for negative inputs, it usually does not cause convergence problems during the training.

If the 0 derivative property of the ReLu function causes difficulties during the training, the so called Leaky ReLu can be used instead. It is defined in (26):

\[ y_j = \max(\alpha \cdot x_j, x_j), \]  \hspace{1cm} (24)

where \( \alpha \) is a user defined parameter, or a learnable free parameter. The derivative of the Leaky ReLu is \( \alpha \) or 1 depending on the input \( x_i \). The ReLu and Leaky Relu functions are illustrated in Figure 10.

These two activation functions are the most commonly used ones in deep neural networks as they stabilize the gradient during the backpropagation and help to avoid the exploding or vanishing gradient problem.
Another useful way to help the training of deep structures is the appropriate initialization of the weights in the network. As in the beginning of the training no information is known about the correct weight values, there is no other option than random initialization. The zero mean value is a natural choice, but the applied variation can influence the training process greatly. Using too large weight values result in large activations and large multiplied gradients, therefore during the training process, the optimization steps will be large and we will not be able to approach the optimal values. On the other hand, when the weight values are small the activations and the multiplied gradient values become small as well, therefore the optimization steps will be very small and it will take a lot of time for the process to converge.

To set the initial weights in an acceptable range two common methods are available, the Xavier [61] and the He [62] methods of initialization. Both approaches generate weights according to a normal distribution. They set the expected value to zero and define the variation in the following way:

\[
\begin{align*}
\mu &= 0, \\
\sigma_{Xav} &= \frac{2}{n_i + n_o}, \\
\sigma_{He} &= \frac{2}{n_i},
\end{align*}
\]

where \(n_i\) and \(n_o\) are the number of inputs and outputs in a particular layer. These approaches try to keep the variance of the total input, \(x_j\), 1, assuming that all the inputs are statistically independent and have zero expected value and 1 variance.

In other words, we would like to set \(Var(x_j) = 1\), assuming that \(Var(x_j) = Var(\sum_i y_i w_{ij}) = n_i Var(w_{ij})\). Therefore we initialize \(w_{ij}\) as \(w_{ij} \sim N \left(0, \frac{1}{n_i}\right)\).

Having these ideas considered one can build up a multi-layered network to solve a classification task. However, if we consider special input data, like images, we can further optimize the network, to obtain better results and achieve higher performance.
5.2.3 Convolutional Neural Networks

In conventional multilayer networks the traditional perceptron model creates a connection between each input and output elements therefore it has a lot of free parameters. This makes the whole network sensitive to overfitting, makes the storage of the network complicated, and it also slows down the training process. Further disadvantage of this structure is that, it does not utilize the spatial correlations in the images.

To overcome these limitations convolutional neural networks have been introduced [63]. These networks contain layers that perform convolutional filtering on the input image. The layers can contain multiple kernels and can have varying kernel resolutions. The output of a convolutional layer is a series of images, each one filtered with a different kernel.

For example if we consider a convolutional layer with 5 different kernels, each one with the resolution 3x3, the layer contains only 45 weights independently from the size of the input image.

It is interesting to mention, that the original idea of weight sharing was introduced in 1986 by Rumelhart, Hinton and Williams [64].

CNNs have been proposed in [65] to classify 2D image data. These types of neural networks consist of several convolutional and sub-sampling layers followed by fully connected layers. An example is shown in Figure 11. As Ravi [66] summarized, the main concepts can be expressed in three points:

- The input image is convolved using several small filters.
- The output of the previous step is sub-sampled.
- The output of the sub-sampling is considered as a new image and the convolution and the sub-sampling process is repeated until high level features are extracted.

In this way low level image features are extracted first, then by combining these low level features higher level image descriptors are obtained in the next layer. Repeating
these steps, the required information is extracted from the input image to perform the classification that is the task of the last fully connected layer. The output of the last layer can be directly used to assign labels for each axial slice and to get the confidence vector of the labeling.

![Figure 1](image.png)

11. Figure. CNN topology example.

In this work AlexNet [3] was used for the classification of the 2D grayscale images. The structure of the network follows the originally introduced topology as it can be seen in Figure 12. As the network was pretrained on the ImageNet database, the originally grayscale input slices were augmented to have three color channels.

The network contains eight layers. The first five are convolutional while the remaining are fully connected. Only the last softmax output is modified to fit the current problem size. The first convolutional layer applies 96 different kernels of size 11x11 pixels with a stride of 4 pixels. The second convolutional layer takes as input the output of the first layer and applies 256 kernels of size 5x5. The outputs of the first two layers are pooled using max-pooling to reduce the size of the intermediate representations and reduce memory usage. The third and fourth convolutional layer has 384 kernels of size 3x3, the fifth has 256 3x3 kernels. The output of the fifth layer is max-pooled again. The fully connected layers contain 4096 neurons each.
The image dataset was split to three nearly equal subsets, such that the training, the cross-validation, and the test sets included 66 000, 55 000, and 60 000 labeled images, respectively. The first set was used to train the classifier, the second was used to monitor the accuracy during the training, and the last one was used to evaluate the CNN model (as well as the post-processed result).

In order to simulate all possible patient positions, and to perform some data augmentation, the axial slices of the training dataset (that included images acquired in supine position only) were rotated 0, 90, 180, and 270 degrees. The training was performed within Nvidia DIGITS framework [67] using the Caffe [68] deep learning toolkit. Stochastic Gradient Descent method was applied as a solver. 30 training epochs were performed. The learning rate was set to 0.01 at the beginning and it was divided by 10 after 10 and 20 epochs. It took 5 hours to fine tune a pretrained CNN using a simple (Nvidia Quadro K4000) GPU.
5.2.4 Initial Classification

Table 10 presents the confusion matrix of the applied CNN network evaluated on the test data set. The columns of the table specify the known ground truth labeling while the rows represent the labeling assigned by the trained system.

For each label the Precision value was calculated as the number of the correctly labeled slices divided by the number of all slices that were associated with that specific label by the CNN. The Recall metric is defined as the number of the correctly labeled slices for a specific label divided by the total number of slices belonging to that specific class according to the ground truth labeling. The overall accuracy (the last cell in the last row) is defined as the ratio of the number of all correctly labeled slices to the total number of slices in all classes.

The overall accuracy of the initial classification was 93.4% and the precision varied between 85.1% and 97.4%. The values in the main diagonal represent the correctly labeled slices. The values directly below or above the main diagonal indicate some minor confusion, usually occurring at the border of the anatomical regions. For example 143 HEAD slices were labeled as CHEST. Other confusions mean serious misclassification errors. This effect indicates the independent processing of the slices, as the labeling process does not incorporate the neighborhood information at this point.

Although the majority of the slices were correctly labeled by the CNN, there are some major artifacts that make the result impractical for direct clinical application. Figure 13 shows two examples from the test set. In the images the central coronal slice of the CT exams are shown and each row of the images represent an axial slice of the corresponding CT exam. In Figure 13(a) the initial classification accuracy is 84.7% and it is free from significant misclassifications while Figure 13(b) is one of the worst initial results (accuracy: 67.8%) and contains examples for all major errors (these errors may caused by the non-supine positioning during the acquisition). The left side of the images show the initial labeling (based on the maximal confidence values) and the confidence values are plot on the right side. The different colors indicate different anatomical regions (yellow - BRAIN, olive - HEAD, yellow-green - CHEST, dark green - ABDOMEN-UPPER, green - ABDOMEN-LOWER, dark magenta - PELVIS-UPPER, red - PELVIS-CENTER, magenta - PELVIS-LOWER, dark blue - THIGH, middle blue - SHIN, blue - FOOT).
<table>
<thead>
<tr>
<th>Pred.\True</th>
<th>FOOT</th>
<th>SHIN</th>
<th>THIGH</th>
<th>PE-LO</th>
<th>PE-CR</th>
<th>PE-UP</th>
<th>AB-LO</th>
<th>AB-UP</th>
<th>CHEST</th>
<th>HEAD</th>
<th>BRAIN</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOOT</td>
<td>1670</td>
<td>94</td>
<td>57</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>9</td>
<td>3</td>
<td>90.40%</td>
</tr>
<tr>
<td>SHIN</td>
<td>64</td>
<td>5662</td>
<td>78</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>97.40%</td>
</tr>
<tr>
<td>THIGH</td>
<td>17</td>
<td>89</td>
<td>5273</td>
<td>318</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>8</td>
<td>15</td>
<td>4</td>
<td>92.10%</td>
</tr>
<tr>
<td>PE-LO</td>
<td>0</td>
<td>0</td>
<td>133</td>
<td>3377</td>
<td>106</td>
<td>4</td>
<td>146</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>89.70%</td>
</tr>
<tr>
<td>PE-CR</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>399</td>
<td>3782</td>
<td>257</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>85.10%</td>
</tr>
<tr>
<td>PE-UP</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>51</td>
<td>518</td>
<td>15140</td>
<td>0</td>
<td>130</td>
<td>34</td>
<td>12</td>
<td>0</td>
<td>95.30%</td>
</tr>
<tr>
<td>AB-LO</td>
<td>0</td>
<td>0</td>
<td>33</td>
<td>61</td>
<td>193</td>
<td>549</td>
<td>14335</td>
<td>801</td>
<td>40</td>
<td>11</td>
<td>0</td>
<td>89.50%</td>
</tr>
<tr>
<td>AB-UP</td>
<td>21</td>
<td>0</td>
<td>15</td>
<td>25</td>
<td>0</td>
<td>11</td>
<td>239</td>
<td>12455</td>
<td>186</td>
<td>45</td>
<td>0</td>
<td>95.80%</td>
</tr>
<tr>
<td>CHEST</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>28</td>
<td>11</td>
<td>265</td>
<td>436</td>
<td>15087</td>
<td>143</td>
<td>17</td>
<td>94.30%</td>
</tr>
<tr>
<td>HEAD</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>18</td>
<td>39</td>
<td>19</td>
<td>67</td>
<td>4033</td>
<td>68</td>
<td>94.80%</td>
</tr>
<tr>
<td>BRAIN</td>
<td>30</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>121</td>
<td>5968</td>
<td>97.40%</td>
<td></td>
</tr>
<tr>
<td>Recall</td>
<td>92.50%</td>
<td>96.90%</td>
<td>94.20%</td>
<td>79.60%</td>
<td>81.70%</td>
<td>94.60%</td>
<td>95.40%</td>
<td>89.90%</td>
<td>97.80%</td>
<td>91.90%</td>
<td>98.40%</td>
<td>93.40%</td>
</tr>
</tbody>
</table>

10. Table. Confusion matrix of the initial labeling.
As the labeling of the slices are independent, the CNN can produce alternating labeling near the region borders (e.g. between PELVIS-CENTER and PELVIS-UPPER) or can completely miss the labeling as can be seen in the HEAD and CHEST regions (see Figure 13(b)). This confusion of non-neighboring regions are major classification errors and affect 1.5% of the slices of the whole test set.

![Figure 13](image)

13. Figure. Initial classification results. (a) Acceptable initial classification with minor errors, (b) Initial classification with major errors.

The above presented initial classification can be also compared with the results of papers [15] and [21]. The referenced results are presented in Tables 5 and 6 respectively. The presented neural classification distinguishes eleven regions while the other two methods use six and five regions only. Considering the increased number of regions, the slightly lower classification accuracy, 93.40% against the 94.0% and 94.1% values, does not mean a notable difference. However, the initial classification is competitive with the other methods, it still produces the anomalies that make the usability of the classification limited in the clinical practice.
5.3 Comparing the Initial Classification Results

In the previous sections, two possible solutions were presented to solve the initial classification. The first method used the MRI data set, traditional image processing and machine-learning methods to solve the task while the second approach has connected the feature extraction and machine learning by using a convolutional neural network.

Considering only the achieved results, it can be said that the two approaches have similar performance. The first version has 91.27% accuracy and the second one has 93.4%. But it also has to be taken into consideration that the neural network used eleven regions and the random forest based approach had to make difference between five regions only. This means that the neural network could differentiate between twice more regions while maintaining its accuracy. Another important aspect is that the average result provided by the neural network is closer to the actual result presented in Figure 13(a) than to Figure 13(b) while the random forest result are more noisy and more similar to Figure 5.

Nevertheless, the most important difference between the two approaches is not in the numbers. From a developer point of view, in the first case, a serious effort has to be taken to find the appropriate image descriptor and machine learning tool to achieve the required accuracy. In case of the neural network, the training process itself finds the correct image features and tunes the network for high accuracy. This means, that a lot of time in the development phase can be spared.

Of course, the usage of the neural networks has its own prerequisites. Significant amount of training data and computation power are required to tune a complex neural network. In this task the amount of available images was high enough to provide a reliable training set and the Nvidia GPUs provided the computational power.

In my opinion, if everything is given to use neural networks, the CNNs can achieve better results while can save a lot of effort for the developer as they can handle simple classification tasks on their own.
6 Establishing Coherent Labeling

Up to this point in the proposed method, the slices are handled independently in the initial classification phase. This approach leads to some errors which make the recognition process less reliable. These errors are corrected by a post-processing step described in this section. This step incorporates some a priori information about the human anatomy in the recognition process to establish a coherent final labeling.

The first and most obvious fact that is exploited is the order of the anatomical regions. The labels shall follow each other in the right order: LEG, ..., LEG, PELVIS, ..., PELVIS, ABDOMEN, ..., ABDOMEN, CHEST, ..., CHEST, HEAD&NECK, ..., in case of the MRI data set. In the CT data set the correct order is FOOT, ..., FOOT, SHIN, ..., SHIN, THIGH, ..., THIGH, PELVIS_LOWER, ..., PELVIS_LOWER, PELVIS_CENTER, ..., PELVIS_CENTER, PELVIS_UPPER, ..., PELVIS_UPPER, ABDOMEN_LOWER, ..., ABDOMEN_LOWER, ABDOMEN_UPPER, ..., ABDOMEN_UPPER, CHEST, ..., CHEST, HEAD, ..., HEAD, BRAIN, ... . These orders cannot be changed, as they are a major property of the human body.

The second consideration is that the size of the anatomy regions cannot be arbitrarily large or small. It can vary among patients but each region has its expected length and variance. The post-processing can incorporate these lengths to estimate the final labeling. The mean size of each region was calculated based on the manually labeled test databases that are large enough to have a good estimation for these statistical values. The mean region sizes are displayed in Table 1 for the MRI data set and the same features are indicated in Table 4 for the CT data set.

The main goal of this step is to produce a continuous and reliable labeling for the whole image series. A labeling can be considered acceptable if it meets the following requirements:

- The labeling must be continuous. No alternating labeling can occur.
- The anatomical regions should appear in their correct order e.g. the CHEST region should be followed by the HEAD
• The size of the anatomical regions should be reasonable. The regions cannot be arbitrary small or large.

To obtain the desired properties of the labeling, a region membership function-fitting algorithm is proposed. In this method the anatomical regions are represented with generalized normal distributions (GDF) [69] as defined in (26), where \( \mu \) is the location, \( \alpha \) is the scale, \( \beta \) is the shape parameter and \( \Gamma(y) \) is the gamma function. Examples of this function can be seen in Figure 14:

\[
GDF(x) = \frac{\beta}{2\alpha \Gamma(1/\beta)} \cdot e^{\left(-\frac{|x-\mu|}{\alpha}\right)^\beta}.
\] (26)

Each membership function can be customized to have region specific parameters. To estimate the position (\( \mu \)) and range of the membership functions (\( \alpha \)), the confidence values, provided by the random forest or by the CNN, can be utilized. The \( \beta \) shape parameter is set to a fixed value (\( \beta = 6 \)) in this work.

In the first step of the method, a set of regions, which are possibly involved in the image, is selected. This provides a list of candidate regions. From this list the anatomically correct sequences are assembled and the most likely one is chosen for further processing.

In the next step, a membership function set is produced to represent the selected region sequence. Based on the confidence values of the initial labeling, the position (\( \beta \)) and range (\( \alpha \)) parameters of the functions are optimized. In this way the functions are fit on the image series and a realistic final labeling can be obtained.
6.1 Selecting the Most Likely Region Combination

To estimate the visible regions in the image, the confidence weighted size of each region $S_i$ is calculated as described in (27):

$$S_i = \sum_{k=1}^{N} \text{conf}(k,i)T_k,$$  \hspace{1cm} (27)

where $N$ is the number of the slices, $\text{conf}(k,i)$ is the confidence value provided by the initial classifier describing the likelihood of that slice $k$ is located in region $i$, and $T_k$ is the physical thickness of slice $k$ ($T_k$ is always available in the DICOM header).

If $S_i$ is greater than a predefined limit, the region is considered visible in the image. In this way the small and misclassified regions can be eliminated from the further processing.

As soon as the visible regions are available, all of the anatomically correct sequences of these regions are assembled. For example, if ABDOMEN, CHEST and HEAD&NECK regions are visible in the MRI scan, the following valid region combinations are generated: (ABDOMEN); (CHEST); (HEAD&NECK); (ABDOMEN, CHEST); (CHEST, HEAD&NECK) and (ABDOMEN, CHEST, HEAD&NECK)

To select the most likely sequence the Summarized Accepted Confidences (SAC) is calculated:

$$Rc_i = \frac{\sum_{k=1}^{N} \text{conf}(k,i) \cdot k}{\sum_{k=1}^{N} \text{conf}(k,i)},$$

$$SAC = \sum_{i=1}^{M} \sum_{k=1}^{N} \text{conf}(k,i)[Rc_{i-1} < k < Rc_{i+1}],$$ \hspace{1cm} (29)

where $M$ is the number of regions in the sequence, $Rc_{i-1}$ and $Rc_{i+1}$ are the confidence weighted center of the previous and next regions, respectively. This means the confidence value $\text{conf}(k,i)$ is accepted only if the position of the $k$-th slice is between the centers of the neighboring regions. This eliminates the most interfering outliers from the SAC calculations. This way the SAC value indicates the summarized confidence
values of the slices that are positioned and labeled correctly according to the processed sequence.

Once all the SAC values are calculated, the region sequence with the highest SAC can be selected as the most likely combination of the candidate regions.

6.2 Fitting the Region Membership Functions

After selecting the most likely region sequence, a set of membership functions are fitted on the confidence values. This process estimates the position and range parameters of the functions to provide a continuous and reliable labeling. Regarding these calculations, the following considerations are taken:

- The membership functions should maximize the correctly covered confidence values.
- The membership functions should minimize the incorrectly covered confidence values.
- The membership functions should minimize the overlap between each other.

To perform the parameter estimation an iterative method is used. The position of the membership functions are initialized with the confidence weighted center of each region and their range property is set to the average size of the represented region. Figure 14(a) shows the initial position and range estimations for a given confidence distribution.

To meet the requirements, a cost function is defined that consists of three components. The first one, defined in (30), penalizes if a membership function, $GDF_j$, does not cover slices that belong to region $j$:

$$cf_1 = \sum_{j=1}^{M} \sum_{k=1}^{N} (1 - GDF_j(k)) \cdot (1 + conf(k, j))^2. \quad (30)$$
The second component (31) penalizes if a membership function covers slices that do not belong to that specific region:

\[ cf_2 = \sum_{j=1}^{M} \sum_{k=1}^{N} \sum_{i=1,i \neq j}^{M} GDF_j(k) \cdot (1 + conf(k,i))^2. \] (31)

The third component of the cost function Eq. (32) tries to minimize the overlap between the region membership functions:

\[ cf_3 = \sum_{j=1}^{M} \sum_{k=1}^{N} \sum_{i=1,i \neq j}^{M} \min \left( GDF_j(k), GDF_i(k) \right). \] (32)

The final cost function is defined as the sum of the three components:

\[ cf = cf_1 + cf_2 + cf_3. \] (33)

To perform the optimization and to find the desired values for the \(\alpha\) and \(\mu\) parameters of the region membership functions a gradient descent iterative optimization method is applied.

As the cost function values highly depend on the input image series and on the performance of the underlying initial classification tool, it is hard to guarantee the smoothness of the function. Therefore, the gradient descent optimization tool can stuck in local optima. However, in practice, the optimization gave reasonable and accurate solutions and no complex global optimization methods had to be used to estimate the model parameters.

Figure 14(b) shows the membership functions after the post-process. It can be observed that the functions are positioned according to the input confidence values and their range fit the confidences as well. The correct ordering of the regions is maintained during the iterations and their size is in the acceptable range.
14. Figure. Membership function fitting. (a) Initialization of the membership functions, (b) Final position and range values of the membership functions.
7 Evaluation

In this chapter the proposed post-process method is evaluated on the two data sets. The initial results, presented in Section 5.1.7 and 5.2.4, are further refined with the region membership fitting process.

7.1 MRI Results

As the MRI dataset contains whole body images only, the evaluation was split into two different parts. In the first part, the method was tested on the original whole body images while in the second part partial body images were acquired from the original ones and the evaluation was performed on the partial images. This step was necessary to obtain a reliable estimation about the real performance of the method, as whole body images are seldom acquired in the clinical practice.

7.1.1 Evaluation on Whole Body MRI Cases

The proposed method was evaluated first on the whole body MRI dataset. The initial classification was performed by the method presented in Section 5.1 and the results were further filtered with the post-process method. The results are presented in Table 11. According to the table the overall accuracy increased to 94.48% and the most important issue of the initial labeling (confusion with non-neighboring region) was completely eliminated. The confusion matrix has nonzero values only in the main diagonal (94.48%) and in cells directly below and above the diagonal (5.52%). This means the majority of the slices are correctly labeled and there are confusions only between neighboring regions. There were no serious mistakes in contrast to Table 9 where 0.72% of the slices were significantly misclassified. As the correct sequence of the labels is always ensured in the result, the confusion between the neighboring regions mean only small displacements of the borders between two regions.
<table>
<thead>
<tr>
<th></th>
<th>true LEG</th>
<th>true PELVIS</th>
<th>true ABDOMEN</th>
<th>true CHEST</th>
<th>true HEAD&amp;NECK</th>
<th>class precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>pred. LEG</td>
<td>1579</td>
<td>92</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>94.49%</td>
</tr>
<tr>
<td>pred. PELVIS</td>
<td>55</td>
<td>1865</td>
<td>101</td>
<td>0</td>
<td>0</td>
<td>92.28%</td>
</tr>
<tr>
<td>pred. ABDOMEN</td>
<td>0</td>
<td>49</td>
<td>1735</td>
<td>46</td>
<td>0</td>
<td>94.81%</td>
</tr>
<tr>
<td>pred. CHEST</td>
<td>0</td>
<td>0</td>
<td>80</td>
<td>1693</td>
<td>10</td>
<td>94.95%</td>
</tr>
<tr>
<td>pred. HEAD&amp;NECK</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>92</td>
<td>2108</td>
<td>95.82%</td>
</tr>
<tr>
<td>class recall</td>
<td>96.63%</td>
<td>92.97%</td>
<td>90.55%</td>
<td>92.46%</td>
<td>99.53%</td>
<td>94.48%</td>
</tr>
</tbody>
</table>

11. Table. Accuracy of slice classification after post-processing using whole body MRI cases.

Figure 15 illustrates the accuracy improvement for one selected case, the same image can be seen in Figure 5b. It shows both the initial classification and the effect of the post-processing method. As it can be seen the most serious misclassifications were corrected, the correct sequence of the labels was established and all the detected regions have a reasonable size and fit the real anatomical structure of the human body. Some further examples are shown in Figure 16. This figure shows results for a few full body MRI cases which demonstrate normal patient variation (concerning sex and level of obesity).
7.1.2 Evaluation on Partial Body MR Images

Since the acquisition of full body scans is not so frequent in the clinical practice (especially in case of MRI modality), the method was tested on partial images as well. The test images were created from the full body scans by using a sliding window sampling. Using this technique, different samples were cut out from the whole-body
scans at different locations with different sample sizes. The cut size varied from 350 mm to 1150 mm (with 100 mm steps), and all possible subseries were generated. This means that all together 36,681 different partial body scans were generated which contained 10,513,282 slices.

The evaluation was performed similarly to the previous cases. The slices were independently labeled with the initial method and then, the post-process was applied, but in these cases, the post-process had significantly less information to work with.

<table>
<thead>
<tr>
<th></th>
<th>true LEG</th>
<th>true PELVIS</th>
<th>true ABDOMEN</th>
<th>true CHEST</th>
<th>true HEAD&amp;NECK</th>
<th>class precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>pred. LEG</td>
<td>943614</td>
<td>88995</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>91.38%</td>
</tr>
<tr>
<td>pred. PELVIS</td>
<td>91020</td>
<td>2447127</td>
<td>118326</td>
<td>0</td>
<td>0</td>
<td>92.12%</td>
</tr>
<tr>
<td>pred. ABDOMEN</td>
<td>0</td>
<td>238429</td>
<td>2921987</td>
<td>60043</td>
<td>0</td>
<td>90.73%</td>
</tr>
<tr>
<td>pred. CHEST</td>
<td>0</td>
<td>0</td>
<td>217707</td>
<td>2195947</td>
<td>20135</td>
<td>90.23%</td>
</tr>
<tr>
<td>pred. HEAD&amp;NECK</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>83181</td>
<td>1086771</td>
<td>92.89%</td>
</tr>
<tr>
<td>class recall</td>
<td>91.20%</td>
<td>88.20%</td>
<td>89.69%</td>
<td>93.88%</td>
<td>98.18%</td>
<td>91.27%</td>
</tr>
</tbody>
</table>

Table 12. Table. Accuracy measured on partial MRI scans.

Table 12 summarizes the accuracy of the anatomy labeling measured on the partial body scans. The numbers show the overall accuracy was 91.27% that is somewhat lower than in case of whole body scans. The table shows that the majority of the slices are correctly labeled, there were confusions only between the neighboring regions (8.73%), and there were no serious misclassifications.

To show a more detailed picture about the accuracy, a graph is presented in Figure 17. The continuous black line represents the average overall accuracy rate, the dotted line shows the minimal, and the dashed line shows the maximal accuracy rate belonging to test images having the same length. The diagram shows that, the worst accuracy was observed at a 350 mm scan which had 40% labeling accuracy, but as it can be seen in Table 12 it does not mean a complete mislabeling, it means only larger displacement of region boundaries. Pay attention also to the high average accuracy that is nearly 90% in
all cut size, which means the majority of the partial scans are correctly labeled. It is also remarkable that the minimal accuracy rises with the increase of the width of the partial scan, and exceeds 90% when the partial scan is longer than 1150 mm. This effect is due to the information content of the initial classification. As the length of the exam is rising, the post-process has more and more information to estimate the correct labeling and to eliminate the false classifications of the initial classification.

17. Figure. Accuracy measure on partial MRI body scans with various size.

To further analyze the accuracy of the results, Table 13 is presented. This table shows the same statistical values seen in Figure 17 extended with the standard deviation of the overall accuracy for each cut size. The low deviation values (<10%) indicate the majority of the results are very close to the average accuracy and only a few mislabeled results belong to the minimal accuracy values.

An accuracy drop can be observed at 750mm in the minimum accuracy curve. This can be explained by the discrete decisions made during the optimization process and the statistical variations in the random forest classification. Considering that the average region size is 215 mm, a partial scan, having 750 mm length, can contain three or four anatomical regions. This discrete decision has to be made in the post-process and can have an impact on the final accuracy. However, this accuracy drop does not influence the tendency as the average accuracy at 750 mm is 91.68%, only the variance is larger (5.21%) compared to the other values. To illustrate the effect of the discrete decisions in post-process Figure 18 is presented. It shows a special case where two cuts slightly differ in position (a few cm). In the left image the leg region is included but the right
image does not contain this region. This few centimeter difference causes a drop in the confidence weighted size of the leg region, therefore it falls below the detection limit in right image.

To illustrate the reliable results on partial scans Figure 19 is presented. It contains examples with different cut sizes and different patient data.

<table>
<thead>
<tr>
<th>Scan width (mm)</th>
<th>Min</th>
<th>Max</th>
<th>Average</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>350</td>
<td>40.49%</td>
<td>100.00%</td>
<td>89.15%</td>
<td>9.09%</td>
</tr>
<tr>
<td>450</td>
<td>45.26%</td>
<td>100.00%</td>
<td>90.79%</td>
<td>5.90%</td>
</tr>
<tr>
<td>550</td>
<td>57.74%</td>
<td>99.69%</td>
<td>89.59%</td>
<td>6.44%</td>
</tr>
<tr>
<td>650</td>
<td>70.68%</td>
<td>99.74%</td>
<td>91.16%</td>
<td>4.68%</td>
</tr>
<tr>
<td>750</td>
<td>62.81%</td>
<td>99.56%</td>
<td>91.68%</td>
<td>5.21%</td>
</tr>
<tr>
<td>850</td>
<td>76.83%</td>
<td>99.20%</td>
<td>92.59%</td>
<td>3.33%</td>
</tr>
<tr>
<td>950</td>
<td>82.07%</td>
<td>99.10%</td>
<td>93.54%</td>
<td>2.88%</td>
</tr>
<tr>
<td>1050</td>
<td>87.03%</td>
<td>99.35%</td>
<td>94.72%</td>
<td>2.55%</td>
</tr>
<tr>
<td>1150</td>
<td>93.64%</td>
<td>99.41%</td>
<td>96.53%</td>
<td>2.09%</td>
</tr>
</tbody>
</table>

13. Table. Accuracy statistics for labeled partial body MR images.

18. Figure. Effect of the estimation of the number of anatomical regions.
7.1.3 Examples of Pathological Cases

In the clinical practice the acquisition of pathological cases is part of the daily examination routines. The method shall be prepared and robust in the recognition of such cases. Figure 20 includes some examples from the data set containing pathological distortions.

At each case the right image shows the confidence values returned by the random forest for the whole-body scan. In the middle a coronal view is presented showing the labeling provided by the post-process considering only the critical region and the third image shows an axial view with the pathological lesion. In Figure 20(a) the right lung is filled with fluid, in (b) there is a tumor in the left lung, and in the (c) image the MRI scan is distorted by an implant next to the spine.

It can be observed that confidence values are stable despite the presence of abnormalities and the whole classification workflow can produce reliable classifications.

19. Figure. Results of the anatomy labeling for partial body MR images

(cut width 350mm top; 650mm bottom).
20. Figure. Pathological MRI cases.
7.2 CT Results

The initial classification of the CT dataset is further refined with post processing method. The artifacts of the initial labeling can be eliminated and a continuous and reliable labeling can be obtained. The post-processed results are summarized in Table 14. As one can see the overall accuracy increased to 94.13%. Furthermore, only the main diagonal and values below and above that are not zero, so the most interfering errors, confusion of non-neighboring regions, were completely eliminated thanks to the post-process. As the post-process guarantees the continuous labeling, the values directly below and above the main diagonal of the confusion matrix mean only small displacements of the borders between the neighboring regions.

To illustrate the effect of the post-process Figure 21(a) is presented. Compared to Figure 13(b) it can be seen, that the initial misclassification errors are eliminated and a continuous labeling is obtained without alternating labels near the borders of the regions. Figure 21(b) shows the ground truth labeling of the examined CT series. It can be observed that the order of the anatomical regions is correct, most of the regions have the correct size and only the PELVIS-LOWER region became smaller than the original one in the ground truth image.
<table>
<thead>
<tr>
<th>Pred.\True</th>
<th>FOOT</th>
<th>SHIN</th>
<th>THIGH</th>
<th>PE-LO</th>
<th>PE-CR</th>
<th>PE-UP</th>
<th>AB-LO</th>
<th>AB-UP</th>
<th>CHEST</th>
<th>HEAD</th>
<th>BRAIN</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOOT</td>
<td>1769</td>
<td>138</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>92.76%</td>
</tr>
<tr>
<td>SHIN</td>
<td>36</td>
<td>5616</td>
<td>113</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>97.42%</td>
</tr>
<tr>
<td>THIGH</td>
<td>0</td>
<td>91</td>
<td>5419</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>91.83%</td>
</tr>
<tr>
<td>PE-LO</td>
<td>0</td>
<td>64</td>
<td>3359</td>
<td>152</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>93.96%</td>
</tr>
<tr>
<td>PE-CR</td>
<td>0</td>
<td>0</td>
<td>495</td>
<td>3942</td>
<td>279</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>83.59%</td>
</tr>
<tr>
<td>PE-UP</td>
<td>0</td>
<td>0</td>
<td>533</td>
<td>15292</td>
<td>214</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>95.34%</td>
</tr>
<tr>
<td>AB-LO</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>428</td>
<td>14505</td>
<td>887</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>91.69%</td>
</tr>
<tr>
<td>AB-UP</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>305</td>
<td>12431</td>
<td>306</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>95.32%</td>
</tr>
<tr>
<td>CHEST</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>531</td>
<td>15088</td>
<td>249</td>
<td>0</td>
<td>95.08%</td>
</tr>
<tr>
<td>HEAD</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>38</td>
<td>4025</td>
<td>89</td>
<td>0</td>
<td>96.94%</td>
</tr>
<tr>
<td>BRAIN</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>115</td>
<td>5978</td>
<td>98.11%</td>
</tr>
<tr>
<td>Recall</td>
<td>98.01%</td>
<td>96.08%</td>
<td>96.84%</td>
<td>79.13%</td>
<td>85.20%</td>
<td>95.58%</td>
<td>96.55%</td>
<td>89.76%</td>
<td>97.77%</td>
<td>91.71%</td>
<td>98.53%</td>
<td>94.13%</td>
</tr>
</tbody>
</table>

Table. Accuracy of slice classification after post-processing using CT cases.
21. Figure. Post-processed CT result. (a) Corrected Labeling. Accuracy: 84.5%, (b) Ground truth labeling.

In Figure 22 other series can be seen to demonstrate the accuracy of the presented method. It can be observed that the method works well in challenging cases. Figure 22(a) shows a patient with high level of obesity and the image has an artifact in the right arm, (b) is a long leg section, in (c) a large tumor can be seen in the left lung, (d) and (e) are partial body scans that are frequently acquired in the clinical practice while (e) shows a scan with non-supine patient position during the acquisition.
22. Figure. Examples for different imaging challenges. (a) High level of obesity and imaging artifact. Accuracy: 94.6%, (b) Partial body scan. Long leg region. Accuracy: 96.7%, (c) Pathological distortion in the lung. Accuracy: 94.4%, (d) Partial body scan: Head and Neck. (e) Partial body scan. Chest. Accuracy: 94.7%, (f) Non-supine position (laying on the right side). Accuracy: 85.4%.
7.3 Reevaluating the MRI Dataset with the Neural Network Based Approach

In the previous paragraphs the original data were filtered with the proposed post-processing method. From those results, it can be seen that the neural network could handle a more complex classification task without a drastic drop in the initial accuracy. Based on this observation I have reevaluated the MRI dataset using the neural approach.

To perform this evaluation I have retrained the network on the MR images. I used the same data split to get the training and the test date set from the original MRI exams. I used the same Alex net topology and meta parameters that were used for the CT cases. During the training process 80% of the training data was used for the actual training and 20% of it was used to monitor the progress. The training was stopped after 63 epochs as the network could not improve its results on the validation data.

After the training session, the network was evaluated on the test exams. The overall accuracy was 93.24%. The confusion matrix is presented in Table 15. The results can be compared to the ones shown in Table 9. One can see that the overall accuracy is slightly better using the neural network, as the original random forest classifier produced only 91.27%. This is not a huge difference but the neural network could outperform the classical approach.

It is interesting to observe the misclassifications in confusion matrix. The neural approach could not avoid serious errors, where the confusion is not between neighboring regions. Around 1.23% of the test slices were seriously misclassified. This means that the output of the network can not be used directly in a clinical application.

<table>
<thead>
<tr>
<th></th>
<th>true LEG</th>
<th>true PELVIS</th>
<th>true ABDOMEN</th>
<th>true CHEST</th>
<th>true HEAD&amp;NECK</th>
<th>class precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>pred. LEG</td>
<td>1609</td>
<td>155</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>91,11%</td>
</tr>
<tr>
<td>pred. PELVIS</td>
<td>6</td>
<td>1659</td>
<td>58</td>
<td>36</td>
<td>0</td>
<td>94,31%</td>
</tr>
<tr>
<td>pred. ABDOMEN</td>
<td>0</td>
<td>142</td>
<td>1832</td>
<td>93</td>
<td>0</td>
<td>88,63%</td>
</tr>
<tr>
<td>pred. CHEST</td>
<td>19</td>
<td>50</td>
<td>16</td>
<td>1671</td>
<td>26</td>
<td>93,77%</td>
</tr>
<tr>
<td>pred. HEAD&amp;NECK</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>30</td>
<td>2091</td>
<td>98,12%</td>
</tr>
<tr>
<td>class recall</td>
<td>98,47%</td>
<td>82,70%</td>
<td>95,62%</td>
<td>91,26%</td>
<td>98,73%</td>
<td>93,24%</td>
</tr>
</tbody>
</table>

15. Table Initial MRI classification results using a neural network.
To eliminate the misclassification, or to minimize their impact at least, the initial results were filtered with the proposed post-processing method. This way, the same region membership functions were fit on the confidence values output of the neural network.

The result of the filtering is presented in Table 16. The process increased the overall accuracy to 94.56% and removed the major classification errors. It can be seen from the confusion matrix, as the values only in the main diagonal, directly above and below it are not zero. As the post-processing guarantees the continuous labelling and the correct order of the anatomical regions, these misclassifications mean only small displacements of the region borders.

<table>
<thead>
<tr>
<th></th>
<th>true LEG</th>
<th>true PELVIS</th>
<th>true ABDOMEN</th>
<th>true CHEST</th>
<th>true HEAD&amp;NECK</th>
<th>class precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>pred. LEG</td>
<td>1631</td>
<td>122</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>93.04%</td>
</tr>
<tr>
<td>pred. PELVIS</td>
<td>4</td>
<td>1767</td>
<td>77</td>
<td>0</td>
<td>0</td>
<td>95.62%</td>
</tr>
<tr>
<td>pred. ABDOMEN</td>
<td>0</td>
<td>117</td>
<td>1803</td>
<td>97</td>
<td>0</td>
<td>89.39%</td>
</tr>
<tr>
<td>pred. CHEST</td>
<td>0</td>
<td>0</td>
<td>36</td>
<td>1701</td>
<td>31</td>
<td>96.21%</td>
</tr>
<tr>
<td>pred. HEAD&amp;NECK</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>33</td>
<td>2087</td>
<td>98.44%</td>
</tr>
<tr>
<td>class recall</td>
<td>99.76%</td>
<td>88.09%</td>
<td>94.10%</td>
<td>92.90%</td>
<td>98.54%</td>
<td>94.56%</td>
</tr>
</tbody>
</table>

16. Table Filtered neural MRI results.

From this reevaluation, it can be seen that the neural workflow can be easily adopted to the MR modality. Its performance in the initial classification is slightly better than the conventional approach’s while the post-processed results are almost the same (94.56% and 94.48% respectively).

It could have been also a great experiment to integrate both modalities into a single network, but the differences in the labelling made the direct integration not possible.
8 Conclusion

In this work I have presented a method with two possible initial classification techniques to solve the problem of the recognition of axial slices in 3D medical images.

The complete method starts with an image preprocessing stage that corrects some translation differences caused by the patient laying and a field of view normalization. This step reduces the variance of the images and makes the work of the following steps easier.

In the next phase, two methods are presented to perform the initial slice-by-slice classification. The first method uses traditional image processing tools. The normalized axial slices are processed with the Zernike transform that creates a lower dimensional representation of the images. This transform uses circular basis function to represent the images. The structure of the basis functions fits well to the human anatomy and it is rotational invariant therefore the representation contains all the required information to solve the initial classification task.

However the Zernike transform creates a smaller representation, this description is still too complex, therefore the PCA transform is applied to them. It makes the representation even more compact while preserving as much information as possible. The obtained representation serves as a good and compact feature vector that can be further analyzed.

To perform the actual classification a Random Forest classifier is used. This machine-learning tool uses multiple decision trees each one trained with a random subset of the training data. As it can be seen from the classification results, it can solve the task with good accuracy.

The second method I have presented solves the classification with a convolutional neural network. It omits all the intermediate steps of the previous method and gives the classification result directly from the input image.

By considering the structure and the working mechanism of the neural network, it is more precise saying that the neural network incorporates the intermediate steps. It also performs a feature extraction and creates a compact representation, but these steps are
not separated. The network learns these transforms during the training phase. It selects the optimal transform, which fits the data best, automatically.

This version of the initial classification performs better in general even in a more complex environment with more regions to recognize.

If we check the results in numbers only, we can see that the achieved accuracy is promising in both methods. However, if some of the MRI or CT scans are examined as a whole, we can see some major classification errors. This can happen because the axial slices of a scan are processed independently. The continuity of the achieved labeling is not guaranteed and no anatomical knowledge is incorporated in the methods to check the validity of the results.

These insufficiencies are corrected in the presented post-processing method. It selects the anatomical regions that are most likely present in the image, based on the confidence values of the initial classification, and fits the regions on the whole scan. During the fitting it maintains the correct order of the anatomical regions, selects their appropriate positions and gives a good estimation about their sizes. This process takes the anatomical knowledge into consideration to ensure a continuous labeling without outliers, too small or too large regions.

The achieved results show that the formed system can solve the classification task reliably. The achieved accuracy is decent and free from the most interfering major artifacts. The methods were tested on an MRI and a CT dataset with five and eleven regions respectively. In both cases, the tests demonstrated the reliability of the system.

Finally, let me share some thoughts about the project. I started this work in 2010 when modern convolutional neural networks were not available. Image classification was a general problem and it was not clear how to solve a specific task. It required a lot of trial and error steps to get an acceptable result. I also tried several methods and failed until I finally found the Zernike transform. It was the breakthrough. Finally, some promising result was achieved. Of course, tuning and parameter optimization had to be done before the presented results were obtained.

I think in the classical process, the correct choice of the image descriptors is the key. It is really hard to tell the computer what can be seen in an image. Sometimes it is just
mathematically hard to describe, sometimes it is even too complex for the developer. For example, a well defined geometry is just mathematically complex, while the concept of the human chest is something that human experts can argue about. Therefore, it is essential to find the right representation for the examined data set.

In this case, the Zernike transform was the solution. It generated a small dimensional representation. It could capture the required information to perform the classification. I think the circular basis functions that are used in the Zernike transform meets the circular structures of the human body and therefore it can generate good feature vectors.

The transform itself generates, of course, just a set of numbers that are really hard to interpret to a human observer but here becomes machine learning handy. It can find statistical relationships in the data sets that remains hidden otherwise. Random forests are strong classifiers, and it turned out to be a good decision to apply them. The results were far better than all the previous non-machine learning based ones that I tried. The Zernike transform and random forest solved the problem.

The results were very promising, but the development process was slow. I was sure that the workflow I followed could not be a general one. It took a lot of time to find an appropriate representation to separate different images and I was sure that my findings could not be reused in other classification tasks.

At this point, I started looking for other methods. My attention turned to neural networks. There were new papers, at that time, which used neural networks to solve similar image classification tasks. The idea seemed general. Configuring a network, assembling a large enough data base and, with a right amount of computational power, the task is solved. I think this idea was mainly right. I had two data sets with labeling and it was not so hard to find a tool that supported convolutional neural networks, so I set up a simple network and let it learn. Undoubtedly, I had to face new problems and I had to find new solutions but I could reproduce my previous results in a few weeks.

Comparing the two methods, from developer point of view, I think the neural networks are more favorable. If a large enough data base and sufficient computing power is available, than using a neural network framework is more comfortable and faster.
compared to the classical image processing methods. In this project both prerequisites were met and accordingly the results were convincing.

In medical image processing however one has to make sure that the methods can work reliably even in unusual conditions. The independent processing of the slices makes the classification error prone. This property comes from the underlying statistical models of the machine learning tools, therefore I would not recommend using them directly in real life clinical situations. The introduced post processing was necessary to ensure anatomical correctness and to make the whole process solid.

As a final conclusion, I found machine learning tools especially useful in image classification, but good data sets and large computational power is needed to make them work. They can find correlations and distinctive properties automatically. This is an important quality, as very often, even the developer cannot describe images and objects accurately. But these methods are not some kind of ultimate problem solvers. They just build statistical models based on the available data and therefore they will produce errors in some situations. The developer has to be aware of this kind of behavior and has to work out methods to detect these situations and to correct the results as much as possible.
9 Acknowledgments

This research work has been carried out at the Department of Control Engineering and Information Technology of the Budapest University of Technology and Economics in collaboration with GE Healthcare Hungary.

I would like to express my thanks to my supervisor, Balázs Csébfalvi for his guidance and support. He taught me how to work effectively as a researcher and I have learned many invaluable lessons from him, which I will always try to follow during my future career.

I would like to thank László Ruskó from the GE Healthcare Hungary for his support. He helped me a lot to be acquainted with medical image processing and to become familiar with its special requirements. He provided also great support during the publications of the results and was always open for new ideas.
10 Thesis points

Thesis 1:

I have defined a workflow to classify the axial slices of CT and MRI images. The workflow consists of three stages. The first step is an image preprocessing stage, that normalizes the axial slices independently for further processing. The second step of the workflow classifies the axial slices and provides information about the class confidence values as well. The third step of the workflow applies a post processing method on the initial classification based on the confidence values. This method incorporates anatomical knowledge to filter unrealistic results from the previous step of the workflow.

Related publications: [J1], [J2], [C1], [C3], [C4], [C5]

Thesis 2:

According to the requirements of the classification workflow, I have developed a method to normalize the input axial images and to provide the initial classification for MRI data. The normalization step performs translation correction, field of view and resolution normalization using classic image processing methods. To generate the initial classification, conventional machine learning tools are used. For image feature extraction the Zernike transform is applied, whose result is further processed with the PCA method to reduce its size. The feature vector is classified by using the Random Forest method, which also returns the classification confidence values.

Related publications: [J1], [C1], [C3], [C4], [C5]

Thesis 3:

I have defined an alternative solution for the image normalization and initial classification. This approach performs translation correction, field of view and resolution normalization for CT images. The initial classification is performed using a convolutional neural network that integrates the feature extraction and the classification process as well. I have used a pretrained AlexNet structure whose last fully connected layer is modified according the needs of the slice classification workflow.

Related publications: [J1], [C2], [C5]
**Thesis 4:**

I have defined a post processing method for the slice classification workflow. This method incorporates anatomical knowledge and can operate on the results of both initial classification methods. It fits region membership functions on the classification confidence values. The fitting method guarantees the correct order of the anatomical regions and estimates the parameters of the membership functions to correctly cover the confidence values.

*Related publications: [J1], [J2], [C1] [C3]*
11 Bibliography


12 Own Related Publications


13 Complete List of Own Publications

<table>
<thead>
<tr>
<th>Number</th>
<th>27</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of publications</td>
<td>27</td>
</tr>
<tr>
<td>Number of journal papers</td>
<td>3</td>
</tr>
<tr>
<td>Number of articles in journals indexed by WoS or Scopus</td>
<td>2</td>
</tr>
<tr>
<td>Number of independent citations:</td>
<td>8</td>
</tr>
</tbody>
</table>

Journal Papers Related to the Theses

[J1] **Tóth, Márton József** ; Ruskó, László ; Csébfalvi, Balázs
Automatic Recognition of Anatomical Regions in Computed Tomography Images

[J2] **Tóth, Márton József** ; Ruskó, László ; Csébfalvi, Balázs
Automatic recognition of anatomical regions in three-dimensional medical images
COMPUTERS IN BIOLOGY AND MEDICINE 76 pp. 120-133. , 14 p. (2016)

Conference Papers Related to the Theses

[C1] **Tóth, Márton József** ; Ruskó, László ; Csébfalvi, Balázs
Változó méretű anatómiai régiók szegmentálása MRI felvételeken

[C2] **Tóth, Márton József** ; Csébfalvi, Balázs ; Ruskó, László
Convolutional Neural Networks for Anatomical Region Detection

[C3] **Tóth, Márton József** ; Ruskó, László ; Csébfalvi, Balázs
Automatic Recognition of Anatomical Regions

[C4] Tóth, Márton József ; Blaskovics, Tamás ; Ruskó, László ; Delso, Gaspar ; Csébfalvi, Balázs
Automated Detection of Anatomical Regions in Magnetic Resonance Images
Jan, Bender; Arjan, Kuijper; Tatiana, von Landesberger; Holger, Theisel; Philipp, Urban (szerk.) Vision, Modeling & Visualization

[C5] Tóth, Márton József ; Ruskó, László ; Csébfalvi, Balázs ; Blaskovics, Tamás
Anatómiai régiók automatikus detektálása
In: KÉPAF 2013 pp. 1-10. , 10 p.

<table>
<thead>
<tr>
<th>Thesis</th>
<th>Related Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[J1], [J2], [C1], [C3], [C4], [C5]</td>
</tr>
<tr>
<td>2</td>
<td>[J1], [C1], [C3], [C4], [C5]</td>
</tr>
<tr>
<td>3</td>
<td>[J1], [C2], [C5]</td>
</tr>
<tr>
<td>4</td>
<td>[J1], [J2], [C1], [C3]</td>
</tr>
</tbody>
</table>

Other Publications, not Related to the Theses

[O1] Gergely, Ferenc Rácz ; Ágota, Kacsó ; Márton, Tóth ; Balázs, Tóth
Fast and accurate initial parameter fitting in Positron Emission Tomography using Neural Networks
Kiss, Bálint; Szirmay-Kalos, László – WAIT 2019 pp. 65-71. , 7 p

[O2] Ágota, Kacsó ; László, Szécsi ; Márton, Tóth ; Balázs, Benyó ; Tamás, Umenhoffe
Finite volume blood flow simulation for highly deformable boundaries
Bálint, Kiss; László, Szirmay-Kalos (szerk.) Proceedings of the Workshop on the Advances of Information Technology: WAIT 2018

[03] Gergely, Ferenc Rácz; Ágota, Kacsó; Márton, Tóth; Balázs, Tóth
PET Image Denoising using a Deep Neural Network
Bálint, Kiss; László, Szirmay-Kalos (szerk.) Proceedings of the Workshop on the Advances of Information Technology: WAIT 2018

[04] Márton, Tóth; Tamás, Umenhoffer; László, Szécsei; Ágota, Kacsó; Balázs, Benyó
Aortic Root Simulation Using Smoothed Particle Hydrodynamics
Bálint, Kiss; László, Szirmay-Kalos (szerk.) Proceedings of the Workshop on the Advances of Information Technology: WAIT 2018

[05] Rácz, Gergely; Kacsó, Ágota; Tóth, Márton; Tóth, Balázs
Enhanced PET Reconstruction with Neural Networks
Szirmay-Kalos, László; Renner, Gábor (szerk.) IX. magyar számítógépes grafika és geometria konferencia, GRAFGE 2018

[06] Tamás, Umenhoffer; Marton, Tóth; Agota, Kacso; Laszlo, Szecsi; Akos, Szlavecz; Peter, Somogyi; Laszlo, Szilagyi; Aniko, Kubovje; Tamas, Szerafin; Laszlo, Szirmay-Kalos et al.
Modeling and simulation framework of aortic valve for hemodynamic evaluation of aortic root replacement surgery outcomes

[07] Tóth, Balázs; Tóth, Márton József; Kacsó, Ágota Enikő; Rácz, Gergely Ferenc; Szirmay-Kalos, László
Controlling TV Regularization with Deep Learning

[08] Umenhoffer, Tamás; Tóth, Márton; Szécsi, László; Kacsó, Ágota; Benyó, Balázs
Aortic Root Simulation Framework for Valve Sparing Aortic Root Replacement Surgery
Bálint, Kiss; László, Szirmay-Kalos (szerk.) Proceedings of the Workshop on the Advances of Information Technology: WAIT 2018

[O9] **Tóth, Márton József**

Progresses in a Fluid Mechanics Simulation Engine on the GPU
Kiss, Bálint; Szirmay-Kalos, László (szerk.) Proceedings of the Workshop on the Advances of Information Technology: WAIT 2017

[O10] **Tóth, Márton József**; Csébfalvi, Balázs

Distribution Interpolation of the Radon Transforms for Shape Transformation of Gray-Scale Images and Volumes

[O11] **Tóth, Márton József**; Csébfalvi, Balázs

Agyszegmens detektálása CT felvételeken inverz anizotróp diffúzióval

[O12] Józsa, Péter ; **Tóth, Márton József** ; Csébfalvi, Balázs

Analytic Isosurface Rendering and Maximum Intensity Projection on the GPU
Vaclav, Skala (szerk.) WSCG 2014 Full Papers Proceedings

[O13] **Márton, József Tóth** ; Balázs, Csébfalvi

Recent Results on Shape-Based Interpolation
Szirmay-Kalos, László; Renner, Gábor (szerk.) VII. Magyar Számítógépes Grafika és Geometria Konferencia

[O14] **Márton, József Tóth** ; Balázs, Csébfalvi

Shape Transformation of Multidimensional Density Functions using Distribution Interpolation of the Radon Transforms
[O15] **Tóth, Márton József**; Csébfalvi, Balázs
Mass-Spring Models for Anisotropic Diffusion
Szirmay-Kalos, László; Renner, Gábor (szerk.) VII. Magyar Számítógépes
Grafika és Geometria Konferencia
Budapest, Magyarország : Neumann János Számítógép-tudományi Társaság

[O16] Török, Marianna; **Tóth, Márton József**; Szöllősi, Alexandra
Foundations and perspectives of mathability in relation to the CogInfoCom
domain
Baranyi, Peter; Esposito, Anna; Niitsuma, Mihoko; Solvang, Bjorn (szerk.) IEEE
4th International Conference on Cognitive Infocommunications: CogInfoCom
2013

[O17] **Márton, Tóth**; Dávid, Dvorszki; Balázs, Csébfalvi
Robust Volume Segmentation using an Abstract Distance Transform
211-217. , 7 p.

[O18] **Márton, Tóth**; Dávid, Dvorszki; Balázs, Csébfalvi
GPU-Accelerated Segmentation of Medical Volume Data
Szirmay-Kalos, László; Renner, Gábor (szerk.) VI. Magyar Számítógépes
Grafika és Geometria Konferencia: GRAFGEO 2012 Budapest, Magyarország:

[O19] Rácz, Gergely Ferenc; Kacsó, Ágota; **Tóth, Márton**; Szirmay-Kalos, László
Dynamic PET Reconstruction from Sinogram Data using Neural Networks
Proceedings of the Workshop on the Advances of Information Technology

[O20] **Tóth, Márton**; Kacsó, Ágota; Magdics, Milán; Salvi, Péter
Advances in the visualization framework for vehicle intelligence simulations
Proceedings of the Workshop on the Advances of Information Technology