

# AUTOMATED DETECTION OF HIGH FDG UPTAKE REGIONS IN CT IMAGES

Annika Liebgott<sup>\*†</sup>    Sergios Gatidis<sup>\*</sup>    Florian Liebgott<sup>†</sup>    Konstantin Nikolaou<sup>\*</sup>    Bin Yang<sup>†</sup>

<sup>\*</sup>Department of Diagnostic and Interventional Radiology, University Hospital of Tübingen, Germany

<sup>†</sup>Institute of Signal Processing and System Theory, University of Stuttgart, Germany

## ABSTRACT

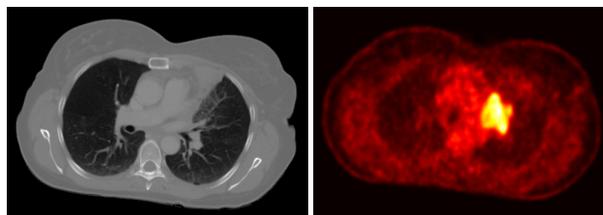
Combined PET-CT scan is an important diagnostic tool in modern medicine, e.g. for staging or treatment planning in the field of oncology. Especially in small structures, like a tumour, textural variations visible in a PET image are not visually recognizable within a CT scan from the same region. Thus, both modalities are necessary for diagnosis. Since both techniques expose the patient to radiation, it would be desirable to get the same information about metabolic activity contained in the PET image from a CT scan only. To investigate the relationship between both imaging modalities, we propose a machine learning approach to automatically identify regions in a CT scan corresponding to areas with high FDG uptakes in a PET image.

**Index Terms**— CT, PET, Radiomics, machine learning, support vector machine

## 1. INTRODUCTION

In modern medicine, medical imaging techniques like Magnetic Resonance Imaging (MRI) or Computational Tomography (CT) are widely-used in clinical diagnostic. While these sophisticated imaging modalities can give detailed visual information about a patient's physiology, they are not able to visualize metabolic activities, a crucial step in some clinical diagnostics. For this purpose, Positron Emission Tomography (PET), which utilizes a positron-emitting tracer injected into the patient's body, is a commonly used method to locate and quantify ongoing metabolic processes [1]. One particularly popular tracer is the radioisotope fluorine-18 which is synthesized into fludeoxyglucose (FDG) and used for the majority of clinical PET scans. In oncology, for instance, FDG-PET is used to observe the regional glucose uptake which corresponds to the concentration of cancer cells and can thus give information about size and location of a tumour, indicate the presence of metastases and help to monitor the success of treatment. The FDG uptake in the affected body tissues is measured in Standardized Uptake Values (SUV). Since PET gives only low resolution information about the patient's physiology beyond the detected glucose uptake, it is often combined with a CT scan to get information about both concentration and detailed location of a tumour [2–4]. Figure 1 presents an example for a CT and a PET image from a PET-CT scan of the thorax. While the CT image clearly shows the anatomic structure of the body, the corresponding PET shows a metabolic activity map.

Although this combination of imaging modalities is a powerful diagnostic tool, one drawback is the high dose of radiation the patient is exposed to. Being able to identify the metabolic activity structures in a CT scan without PET imaging and its radioactive tracer is thus an interesting problem. It is, however, also a challenging task since PET FDG uptake and CT attenuation values are based on different physical and physiological processes. Thus, a direct relation between the two values is not necessarily given. However, the same



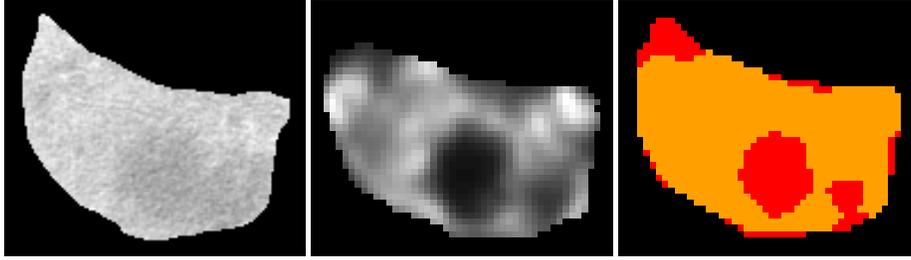
**Fig. 1:** Example of a CT (left) and PET (right) image from a PET-CT thorax scan. Yellow represents a region of high FDG uptake.

physiological processes are the underlying cause for findings in both modalities. For example, tumor necrosis results in reduced FDG uptake but also reduced CT density due to decrease in perfusion. Thus, CT density can be a surrogate for tumor glucose metabolism in such a case.

It is, however, very difficult to find visual indicators for varying SUVs in the CT image (see Figure 2). One possible solution might be found in the field of Radiomics, i.e. the utilisation of machine learning on medical imaging problems. A trained classifier might be able to learn from examples how to detect PET-like image structures in CT which are hardly visible to a human observer. We propose to train a Support Vector Machine (SVM) with patched CT images which are labeled according to the regional FDG uptake in the co-registered PET images from a PET-CT scan. By using this classifier, we then try to locate regions of SUVs higher than a defined threshold.

To the best of our knowledge, there has been no previous work on this specific task. Research in related fields focuses on other problems, e.g. the automated detection or segmentation of tumours by utilizing PET or PET/CT features. Bi et al. [5], for example, conducted a study on detecting regions of high FDG uptake in whole-body-MRI with an SVM trained with features extracted from PET and CT images. Another work presented by Arimura et al. [6] focused on automated contouring of tumours, also by training an SVM with metrics derived from both PET and CT data. Kerhet et al. [7] used features extracted from PET images to train a classifier to predict the most appropriate SUV threshold for tumour segmentation. All works mentioned above use features derived from FDG uptake to classify physiological structures as malignant or benign, rather than analysing the distribution of SUVs within a tumour.

Other research groups focus on intratumour heterogeneity as a method to estimate treatment success or patient survival rates [8–10]. Such studies use features derived from PET or PET/CT images to find links between regional FDG uptake within tumours and patient outcome. They focus more on the relationship between image characteristics and treatment success by either investigating PET features or combining PET and CT, but not on the relationship between both modalities like we do.



**Fig. 2:** A 2D slice from a segmented tumour to demonstrate the visual differences between both imaging modalities. From left to right: CT image, PET image and the label representation of the PET image.

## 2. GENERATING LABELS FROM PET IMAGES

The main objective of this study is to investigate, whether it is possible to identify regions in a CT scan corresponding to areas with high FDG uptakes in a PET image. The focus lies on smaller volumes of interest (VOIs), i.e. already segmented tumour masses. Figure 2 shows an example of a 2D slice of a segmented tumour from a CT scan, the corresponding PET image and the PET images' label representation which indicates the location of high SUV regions compared to the rest of the tumour. Comparing the structures of the PET and CT images, the challenge of visually identifying the PET intensities within the CT scan becomes apparent. The PET image shows distinctive intensity variances, high intensity regions (see orange areas on the label representation on the right) can easily be distinguished from areas of lower SUVs. In the CT image, however, it is hard to interpret gray values in relationship to the FDG-uptake in the PET image. While one lower intensity region located in the lower middle part of the region of interest (ROI) seems to correspond to a slightly darker shade in the same region of the CT, the lower SUV area in the upper left of the PET image cannot be visually identified in the CT scan. It is our hope, that such structures in a CT scan almost hidden to human observers can be detected by a classifier, then amplified and made visible without PET imaging.

For this purpose, we train a classifier with labeled CT images whose labels are derived from co-registered PET images. Both PET and CT images of the VOIs are processed slicewise. Each slice is divided into overlapping patches and the mean intensity of a PET patch is converted into a SUV class label for the corresponding CT patch. These labeled CT patches are used as input to train a classifier that is hopefully able to detect which regions of a new CT image correspond to high SUV regions in a PET image.

First, the typical SUV range of PET images is divided into  $N_c$  intensity bins corresponding to  $N_c$  SUV classes. For each PET patch of size  $p_{PET} \times p_{PET}$ , the mean intensity value is calculated and its corresponding SUV class is used as a label for the respective CT patch of size  $p_{CT} \times p_{CT}$ . Since PET has a lower resolution than CT,  $p_{CT} > p_{PET}$  applies. As a first step, this study focuses on a binary classification task. The patches are thus divided into only two classes: one for high SUVs and one for the remaining patches.

## 3. CLASSIFICATION

### 3.1. Feature extraction

First, features are extracted from the input data, i.e. a labeled CT patch, and stacked into a feature vector  $\underline{x}$ . For this purpose, we used ImFEATbox, our comprehensive toolbox for feature extraction and analysis [11, 12]. We tested several combinations of features based both on visual image characteristics (e.g. texture, contrast) and other

numeric metrics (e.g. moments, transforms). The best results were obtained using a combination of the following categories of features with number of features  $N_f$  per category:

1. Intensity features ( $N_f = 7$ ) [13]
2. Zernike moment features ( $N_f = 92$ ) [14, 15]
3. Fourier transformation features ( $N_f = 300$ ) [12]
4. Gabor filter features ( $N_f = 3600$ ) [16]
5. Lacunarity features ( $N_f = 6$ ) [17]

### 3.2. Feature selection

Before the classification step, the dimensionality of the feature vector  $\underline{x}$  is reduced to obtain a reduced feature vector  $\underline{x}$  for further processing. While reducing the dimension of the feature space via principal component analysis [18] worked quite fast, it was not able to clearly identify a good feature combination for our problem. At the end, 140 principal components were required to achieve an adequate test accuracy.

Sequential floating forward selection (SFFS) [19] proved to be a better option for our problem, despite the high computational cost. With SFFS, the classifier could be trained with only 7 features while still achieving a high test accuracy. Especially features from the families of Fourier transform and Gabor filters were among the first features chosen by SFFS.

Two-dimensional Fourier transform of the CT patches showed a difference in the spectrum between regions with high SUVs and lower SUVs. For all patches, the main energy lies around the zero frequency. The percentage of the energy in the higher frequency bins, however, is roughly 20 times higher in high SUV than lower SUV regions. This confirms our hypothesis, that high SUV regions, although hardly visible to a human observer, can be identified in CT images using suitable features derived from hardly visible structures or transforms of the CT patches.

### 3.3. Classifier

For classification, we chose a binary soft-margin Support Vector Machine (SVM)

$$\hat{y}_i = \text{sgn} \left( \sum_{j=1}^N \alpha_j y_j k(\underline{x}_j, \underline{x}_i) + b \right) \quad (1)$$

$$\text{s.t. } 0 \leq \alpha_i \leq C \quad \forall i = 1, \dots, N$$

$$\text{and } \sum_{i=1}^N \alpha_i y_i = 0$$

where  $\hat{y}_i$  represents the estimated label for sample  $\underline{x}_i$ ,  $\alpha_i$  are the dual Lagrange coefficients,  $C$  a constant soft margin weight and  $b$

the offset to the origin of the coordinate system. For the kernel we use a radial basis function (RBF)

$$k(\underline{x}_j, \underline{x}_i) = \exp(-\gamma \|\underline{x}_j - \underline{x}_i\|^2) \quad (2)$$

where  $\gamma$  is a positive constant.

Tests with a grid search and 5-fold cross-validation to determine the soft-margin parameter  $C$  and the kernel parameter  $\gamma$  showed that a combination of  $C = 2^3$  and  $\gamma = 2^{-15}$  leads to good results. For the implementation of the classifier, we utilised the LIBSVM library [20].

#### 4. EXPERIMENTS AND RESULTS

This study and the conducted experiments, which are described in the following, are intended to be a first proof of concept to investigate whether detecting PET-like image structures within CT images is possible.

##### 4.1. Dataset and experimental setup

We used a total of 41 FDG-PET/CT images from 41 patients with lung cancer, which were acquired on a clinical PET/CT scanner (Siemens Biograph mCT). The segmentation of the primary lung tumour on the CT images to define the volumes of interests (VOIs) was performed by an experienced radiologist. These VOIs were then transferred from CT to the co-registered PET images. For further processing, only image content within the VOIs, i.e. the tumour, was considered (example see Figure 2).

PET and CT images were divided into patches as described in section 2 with patch sizes of  $4 \times 4$  for PET and  $8 \times 8$  for CT, respectively. Intensity values of the PET patches were categorized into  $N_c = 2$  bins (with the threshold at  $2.5 \cdot 10^4$ ) and used as class labels for the corresponding CT patches. This process resulted in a total number of 261032 labeled CT patches for the training and testing of the SVM. The large amount of available training data leads to a high computational complexity for training the classifier and especially for conducting the SFFS. Therefore, since initial tests with a small feature set showed similar results for a reduced number of training samples and the full dataset, we only used the images of 12 patients as dataset  $\mathcal{D}$  for training and test. The images of the remaining 29 patients were later used for validation on unseen patients with varying tumour physiologies to study the generalization capability of the classifier.

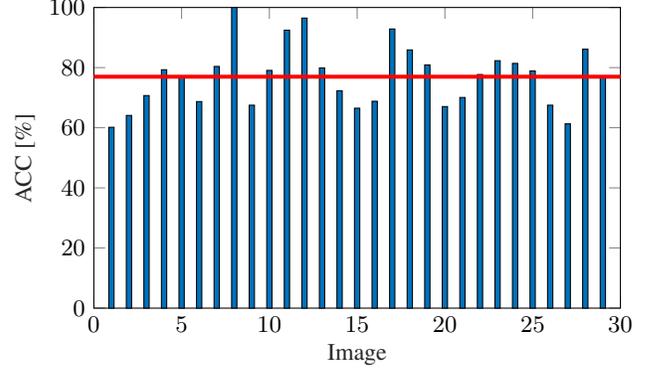
For the training process,  $\mathcal{D}$  was randomly split into a training set and a test set. 70% of the labeled samples were used for training, the remaining 30% for testing. To achieve robust results, the random splitting of the dataset and the training of the SVM was performed 10 times. As a final result, the mean test accuracy of all 10 trained SVMs was calculated

$$\overline{\text{ACC}} = \frac{1}{10} \sum_{i=1}^{10} \frac{\text{TP}_i + \text{TN}_i}{N} \quad (3)$$

where  $\text{TP}_i$  and  $\text{TN}_i$  are the numbers of true positives and true negatives for the  $i$ -th splitting and  $N$  denotes the number of samples under test.

##### 4.2. Classification results

With the 7 features selected by SFFS from our best combination of feature categories (see section 3.2), we achieved a mean test accuracy of  $\overline{\text{ACC}} = 70.94\%$  with a standard deviation of 0.34. While this accuracy still leaves room for improvement, it is a first proof that



**Fig. 3:** Accuracies of all 29 tested images of unseen patients (blue bars) and the mean test accuracy  $\overline{\text{ACC}}$  (red).

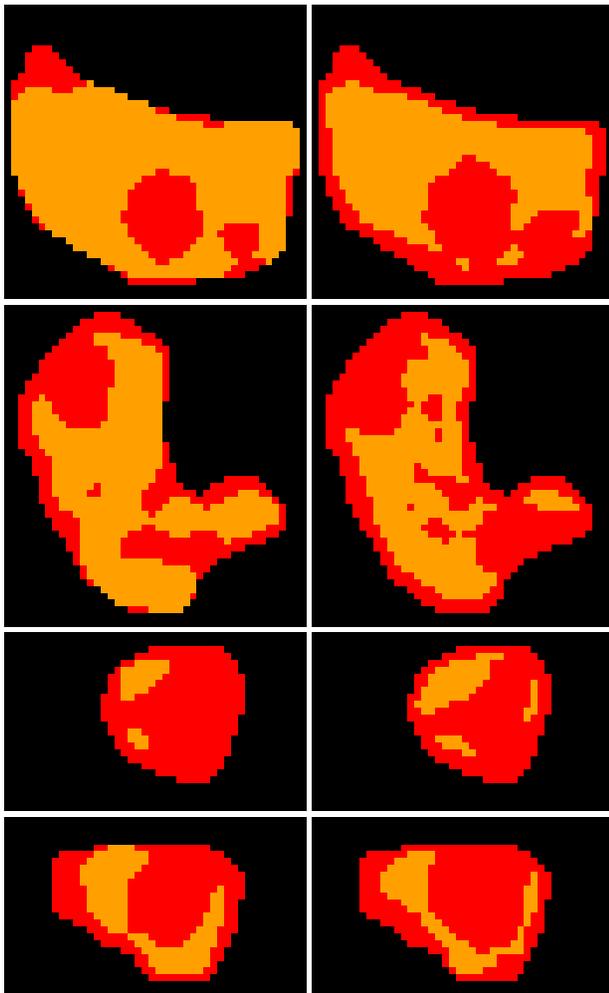
the classifier is generally able to detect PET-like structures within CT images.

We also investigated the ability of the trained SVM to classify CT images of the remaining 29 unseen patients. We achieved a mean test accuracy of  $\overline{\text{ACC}} = 76.97\%$  with a standard deviation of 10.33. Figure 3 shows the results for all unseen patients. The higher accuracy compared to the accuracy achieved with dataset  $\mathcal{D}$  originates in the different nature of the patients' tumours. For example, the image for which an accuracy of 100% was achieved contained no high SUV region, which was correctly perceived by the classifier. The classifier proved to be able to identify most of the regions of high SUVs with a tendency to detect larger or more high SUV areas than are actually present. For large ROIs, the results were considerably better than for small ROIs with only a few or no high SUV regions, e.g. slices from small tumours or the lowest or highest slices of larger tumours. In these smaller ROIs, the SVM tends to have a high false-positive rate, i.e. wrong classification of low SUV as high SUV patches. These findings will be included in the design of future experiments.

The most important aspect in our study is, however, the visual resemblance between the original PET images (converted to their label representation) and the metabolic activity maps estimated from the CT patches. Evaluation of the inhomogeneities within the tumour is an important aspect in e.g. treatment planning in oncology. Thus, it is more important to get a good visual approximation of the high SUV regions than to only look at the absolute accuracy for all patches. Accordingly, we investigated how similar the images actually look. While some images showed a significant amount of errors, there were a lot of images where the classifier was able to correctly identify regions in a CT scan corresponding to areas with high FDG uptakes in a PET image.

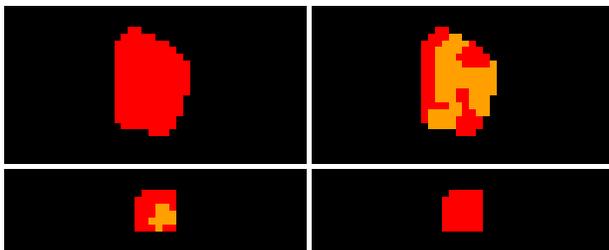
Some examples of 2D slices with well classified high SUV regions are presented in Figure 4. The orange areas represent the high SUV regions and the red areas the remaining tumour. While some patches are wrongly classified, the general shapes of the high SUV regions in the estimated images agree well with these in the PET images, even for difficult shapes like in the images of the two upper-most rows.

Figure 5 shows two examples of images for which the classifier was not able to correctly identify regions of high SUVs. The upper row displays a slice of a tumour, where no regions of high SUVs are present in the original image. The SVM, however, classified more than half of the CT patches as high SUV regions. The opposite is the case for the example in the lower row. It should be noted that both images in this Figure are relatively small compared to the images in Figure 4. This further encourages the assumption that the classifier



(a) original image (b) estimated image

**Fig. 4:** Examples of 2D slices reconstructed from well estimated patch labels (right column) and the corresponding original PET slices converted to labels (left column).



(a) original image (b) estimated image

**Fig. 5:** Examples of estimated and reconstructed 2D slices (right column), where the trained model could not sufficiently determine the right class, compared to the original label images (left column). The upper row shows a case, where a significant amount of high intensity areas were detected while none are present in the original image. In the lower row is an example, where the classifier failed to recognize the high SUVs.

has difficulties to handle patches from smaller structures. The reason might be that patches from such images are influenced too much by the borders between tumour and background.

Both Figures represent an observation we could make for most images: for larger structures, the similarity between original label image and estimated metabolic activity map tends to be higher, for smaller ROIs the accuracy decreases. Other images lie, of course, between both extremes, but the overall tendency suggests that estimating PET-like structures from CT images is generally a feasible problem.

## 5. DISCUSSION

Our experiments showed promising results as a proof of concept as well as some drawbacks of the proposed approach in its current status. We got good visual resemblance between the estimated metabolic activity maps and the original PET label images for larger ROIs. The classifier was able to detect not only the location of many high SUV regions within such ROIs, but also its general shape.

However, our proposed method tends to detect more regions of high SUVs than present. In some cases, especially in smaller tumours or slices on the edges of a VOI, the classifier has difficulties to correctly differentiate between classes. Furthermore, it is up to now only able to perform a binary SUV classification. A more detailed detection of PET structures within the CT images in terms of a multi-class SUV classification is desirable in the future, e.g. to further differentiate between medium and low SUV regions.

To overcome these difficulties, we plan on one hand to investigate other features or feature combinations to find a better representation of the CT images. On the other hand, we are going to experiment with deep neural networks to replace the SVM. This way, we can avoid hand-crafted features and instead let the classifier learn the best representation by himself. Moreover, we intend to extend the classifier to a multi-class model to be able to discriminate between more than two SUV classes.

## 6. CONCLUSION

In this study, we presented a machine learning approach to detect high intensity PET structures within CT images. We were able to show that our approach of using PET intensities as class labels to classify CT image patches works in general. In our experiments, we identified features derived from the Fourier transform and Gabor filters as promising descriptors for identifying regions of high SUVs in CT images. Moreover, we were able to train a classifier to detect such regions, although the results obtained so far leave room for improvements. One major drawback of the proposed method is the decrease in classifier precision for smaller ROIs. Another drawback is the tendency of the classifier to detect too many patches as region of high SUVs.

In the near future, we aim to improve our proposed approach by further investigating which features are most suitable for our application as well as by using a deep learning approach. We also intend to not only be able to identify regions of high SUVs, but to discriminate between more intensity classes to produce a more accurate approximation of the structures of PET images.

## 7. REFERENCES

- [1] R. Edward Coleman, "Clinical PET in oncology," *Clinical Positron Imaging*, vol. 1, no. 1, pp. 15 – 30, 1998.
- [2] Isaac R Francis, Richard K J Brown, and Anca M Avram, "The clinical role of CT/PET in oncology: an update," *Cancer Imaging*, vol. 5, pp. S68 – S75, 2005.
- [3] Andrea Gallamini, Colette Zwarthoed, and Anna Borra, "Positron emission tomography (PET) in oncology," *Cancers*, vol. 6, no. 4, pp. 1821–1889, 2014.
- [4] Medhat M. Osman Nghi C. Nguyen, Isin Akduman, "F-18 FDG-PET and PET/CT imaging of cancer patients," *Journal of Radiology Nursing*, vol. 27, no. 2, pp. 61 – 69, 2008.
- [5] L. Bi, J. Kim, D. Dagan Feng, and M. Fulham, "Classification of thresholded regions based on selective use of PET, CT and PET-CT image features," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug 2014, pp. 1913–1916.
- [6] H. Arimura, Z. Jin, Y. Shioyama, K. Nakamura, T. Magome, and M. Sasaki, "Automated method for extraction of lung tumors using a machine learning classifier with knowledge of radiation oncologists on data sets of planning CT and FDG-PET/CT images," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, July 2013, pp. 2988–2991.
- [7] A. Kerhet, C. Small, H. Quon, T. Riauka, L. Schrader, R. Greiner, D. Yee, A. McEwan, and W. Roa, "Application of machine learning methodology for PET-based definition of lung cancer," in *Current oncology*, 2010.
- [8] Thida Win, Kenneth A. Miles, Sam M. Janes, Balaji Ganeshan, Manu Shastry, Raymondo Endozo, Marie Meagher, Robert I. Shortman, Simon Wan, Irfan Kayani, Peter J. Ell, and Ashley M. Groves, "Tumor heterogeneity and permeability as measured on the ct component of pet/ct predict survival in patients with non-small cell lung cancer," *Clinical Cancer Research*, vol. 19, no. 13, pp. 3591–3599, 2013.
- [9] Sugama Chicklore, Vicky Goh, Musib Siddique, Arunabha Roy, Paul K. Marsden, and Gary J. R. Cook, "Quantifying tumour heterogeneity in 18f-fdg pet/ct imaging by texture analysis," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 40, no. 1, pp. 133–140, Jan 2013.
- [10] I. El Naqa, P. W. Grigsby, A. Apte, E. Kidd, E. Donnelly, D. Khullar, S. Chaudhari, D. Yang, M. Schmitt, Richard Laforest, W. L. Thorstad, and J. O. Deasy, "Exploring feature-based approaches in pet images for predicting cancer treatment outcomes," *Pattern Recogn.*, vol. 42, no. 6, pp. 1162–1171, June 2009.
- [11] Annika Liebgott, Sergios Gatidis, Petros Martirosian, Fritz Schick, Bin Yang, and Thomas Küstner, "ImFEATbox: An MR image processing toolbox for extracting and analyzing features," in *Proceedings of the International Society for Magnetic Resonance in Medicine (ISMRM)*, Honolulu, Hawaii, USA, Apr. 2017.
- [12] "ImFEATbox," <https://github.com/annikaliebgott/ImFEATbox>, Oct. 2017.
- [13] Kiaran P. McGee, Armando Manduca, Joel P. Felmlee, Stephen J. Riederer, and Richard L. Ehman, "Image metric-based correction (autocorrection) of motion effects: Analysis of image metrics," *Journal of Magnetic Resonance Imaging*, vol. 11, no. 2, pp. 174–181, 2000.
- [14] v. F. Zernike, "Beugungstheorie des Schneidenverfahrens und seiner verbesserten Form, der Phasenkontrastmethode," *Physica*, vol. 1, pp. 689–704, May 1934.
- [15] Amir Tahmasbi, Fatemeh Saki, and Shahriar B. Shokouhi, "Classification of benign and malignant masses based on Zernike moments," *Comput. Biol. Med.*, vol. 41, no. 8, pp. 726–735, Aug. 2011.
- [16] Joni-Kristian Kämäräinen, "Gabor features in image analysis," in *3rd International Conference on Image Processing Theory, Tools and Applications, IPTA 2012, October 15 - 18, Istanbul, Turkey*. 2012, pp. 1–2, Institute of Electrical and Electronics Engineers IEEE.
- [17] T.G. Smith Jr., G.D. Lange, and W.B. Marks, "Fractal methods and results in cellular morphology dimensions, lacunarity and multifractals," *Journal of Neuroscience Methods*, vol. 69, no. 2, pp. 123 – 136, 1996.
- [18] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *J. Educ. Psych.*, vol. 24, 1933.
- [19] P. Pudil, F. J. Ferri, J. Novovicova, and J. Kittler, "Floating search methods for feature selection with nonmonotonic criterion functions," in *Proceedings of the 12th IAPR International Conference on Pattern Recognition, Vol. 3 - Conference C: Signal Processing (Cat. No.94CH3440-5)*, Oct 1994, vol. 2, pp. 279–283 vol.2.
- [20] C. Chang and C. Lin, "LIBSVM: A library for support vector machines," *T. Intell. System. Tech.*, vol. 2, pp. 1–27, 2011.