In Y. Zheng, B.M. Williams, K. Chen (Eds.): Medical Image Understanding and Analysis. Communications in Computer and Information Science, Vol. 1065, 38-47, Springer, 2020.
The final publication is available at link.springer.com

Wilms' Tumor in Childhood: Can Pattern Recognition Help for Classification?

Sabine Müller[1,2,*], Joachim Weickert[2], Norbert Graf[1]

[1]Department of Pediatric Oncology and Hematology,
Saarland University Medical Center, Homburg, Germany
[2]Mathematical Image Analysis Group,
Saarland University, Campus E1.7, Saarbrücken, Germany
{smueller,weickert}@mia.uni-saarland.de
graf@uks.eu

Abstract. Wilms' tumor or nephroblastoma is a kidney tumor and the most common renal malignancy in childhood. Clinicians assume that these tumors develop from embryonic renal precursor cells - sometimes via nephrogenic rests or nephroblastomatosis. In Europe, chemotherapy is carried out prior to surgery, which downstages the tumor. This results in various pathological subtypes with differences in their prognosis and treatment.
First, we demonstrate that the classical distinction between nephroblastoma and its precursor lesion is error prone with an accuracy of 0.824. We tackle this issue with appropriate texture features and improve the classification accuracy to 0.932.
Second, we are the first to predict the development of nephroblastoma under chemotherapy. We use a bag of visual model and show that visual clues are present that help to approximate the developing subtype.
Last but not least, we provide our data set of 54 kidneys with nephroblastomatosis in conjunction with 148 Wilms' tumors.

1 Introduction

Wilms' tumor, or nephroblastoma, accounts for 5 % of all cancers in childhood and constitutes the most frequent malignant kidney tumor in children and juveniles [16]. About 75% of all patients are younger than five years - with a peak between two and three years [5,11]. Nephroblastoma is a solid tumor, consisting mainly of three types of tissue: blastema, epithelium and stroma [21]. In Europe, diagnosis and therapy follow the guidelines of the International Society of Pediatric Oncology (SIOP) [6,10]. One of the most important characteristics

of this therapy protocol is a preoperative chemotherapy. During this therapy, the tumor tissue changes, and a total of nine different subtypes can develop [6]. Depending on this and the local stage, the patient is categorized into one of the three risk groups (low-, intermediate-, or high-risk patients) and further therapy is adapted accordingly. Of course, it would be of decisive importance for therapy and treatment planning to determine the corresponding subtype as early as possible. It is currently not known how this can be achieved.
However, there are very few research results in this direction so far: To the best our knowledge, there is only the recent work of Hötker et al. [9], where they show that diffusion-weighted MRI might be helpful in making this distinction. Unfortunately, diffusion-weighted MR images are not yet recorded as standard. Due to a relatively low incidence of this disease, it is also difficult to sensitise the clinical staff in this direction. On the other hand, a T_2 sequence is part of the therapy protocol and always recorded - even if there are no parameter specifications. We can show that even this standard sequence might be sufficient to predict subtype tendencies.
In about 40% of all children with nephroblastoma, so-called nephrogenic rests can be detected. Since these only occur in 0.6% of all childhood autopsies, they are considered a premalignant lesion of Wilms' tumors [2]. The diffuse or multifocal appearance of nephrogenic rests is called nephroblastomatosis [13,15]. Despite the histological similarity, nephroblastomatosis does not seem to have any invasive or metastatic tendencies. In order to adapt the therapy accordingly and not to expose children to an unnecessary medical burden on the one hand and to maximize their chances of survival on the other, it is necessary to distinguish nephroblastoma and its precursor nephroblastomatosis at the beginning of treatment. Its visual appearance has been described as homogenous and small abdominal mass [4,18]. However, all existing publications describe the visual appearance on usually very small data sets [7,18]. So far, it has never been validated statistically to what extent the described features are sufficient for classification. Thus, we review this current clinical practise. For this purpose, we have created a data set and evaluate whether the assumed properties can solve the classification problem between these two entities. In addition, we propose further properties that dramatically simplify the problem.

In summary our main contributions are:

- We demonstrate that the assumptions about nephroblastomatosis are mostly correct, but not sufficient to ensure a reliable classification. We solve this problem by including more texture features in the classification procedure.
- We are the first to show that T_2 imaging can be used to predict tumor development under chemotherapy in advance. We extract a variety of features and create a collection of visual properties from each image. We use this visual vocabulary to create a histogram of the relative frequency of each pattern in an image of a given subtype. We then use this information for subtype determination.

Fig. 1. Exemplary images from our data set. From left to right: epithelial dominant, stromal dominant, blastemal dominant, regressive, nephroblastomatosis.

– We provide a data set with images of nephroblastomatosis and nephroblastoma from a total of 202 different patients.[1]

2 Materials and Methods

Nephroblastoma is the most common kidney tumor in childhood, although it is always difficult to collect a sufficient amount of data from children and adolescents. This problem is partially solved within large-scale multi-center studies on Wilms' tumor [17,22].

2.1 Data Sets

In recent years, the SIOP studies have collected clinical and imaging data from more than 1000 patients, possible through networking of many hospitals. Unfortunately, this has also caused a major problem: The MR images were taken on devices from different manufacturers with different magnetic field strengths over several years. In addition, there are no uniform parameter sets and the individual sequences (of the same type) can vary dramatically.

We made sure that the main parameter settings of the T_2 sequences included in our data set are as similar as possible - this has drastically reduced the amount of imaging data available. Nevertheless, we have compiled a data set of 202 patients, see Tab. 1. All data sets are T_2-weighted images (axial 2D acquisition) with 3.4 mm to 9.6 mm slice thickness and inslice-sampling ranging from 0.3 mm to 1.8 mm.

In a first step, we cubically resampled all images to a grid size of one in x and y direction, but refrained from resampling in z direction as the interpolation error would be too high. Then, we linearly rescaled image intensities for simplicity to the interval $[0, 1]$. In the end, a human expert with years of experience in the field of nephroblastoma annotated the tumor regions using the method of Müller et al. [14]. We mask everything except the tumor areas and embed them in a square shaped image; see Fig. 1.

[1] The data set can be accessed at www.mia.uni-saarland.de/nephroblastomatosis

Table 1. Detailed information about our data set.

Patient characteristics		
Age	range (month)	$1 - 153$
	average	34.3
Gender	female	50.9%
	male	49.1%
Metastasis (Wilms' Tumor)		22 (14.86%)
Tumor characteristics		
Nephroblastoma Subtypes	diffuse anaplastic	3
	blastemal	18
	regressive	50
	mixed	29
	stromal	28
	epithelial	17
	necrotic	3
	total	148
Nephroblastomatosis		54
Total		202

Research Ethics of the Study All images were received as part of the SIOP 2001 prospective clinical trial. This trial received Ethical Approval from 'Ärztekammer des Saarlandes', Germany, No.: 248/13. Informed consent was given by parents or legal guardians of all enrolled children with nephroblastoma. In addition, all DICOM files were anonymized before analysis.

2.2 Features

First, we like to investigate and improve the currently clinically applied distinction between nephroblastoma and nephroblastomatosis. In order to imitate the clinically used properties as accurately as possible, we apply texture features and evaluate their significance.
Next, we evaluate if it is possible to predict the development of a nephroblastoma under chemotherapy based on standard T_2 sequences. Also in this case we like to know if the overall structure of a tumor layer already contains information about the subtype. For this purpose we use a Bag of Visual Words model.

Texture Features Haralick et al. [8] established the basic assumption that gray-level co-occurence matrices contain all available textural information of an image. These second order Haralick texture features are extensively used in recent years in the area of medical image analysis to diagnose and differentiate cancer [20,23,24].
The basis of co-occurrence characteristics is the second-order conditional probability density function of an given image. Here, the elements of the co-occurrence

matrix for the structure of interest represent the number of times intensity levels occur in neighboring pixels. Several features can be extracted from this matrix, e.g. contrast, homogeneity, entropy, autocorrelation [8,19]. We use these features to distinguish nephroblastomatosis and Wilms' tumors.

Bag of Visual Words Model The basic idea of a bag of visual words model is to represent an image as a set of local visual features. For this purpose we calculate the SURF features [1] of each 8th tumor pixel for a patch of size 7×7. The patches of the training images are then clustered with k-means [12] where cluster centroids are visual dictionary vocabularies. This allows us to determine a frequency histogram of the features in each training and test image. We use this information to train a bagged random forest classifier with 300 decision trees [3].

3 Experiments

We use our data set consisting of nephroblastomatoses and Wilms' tumors to perform several experiments. First, we want to know how accurate the clinical assumption is that nephroblastomatosis and Wilms' tumor can be distinguished by size and homogeneity. Then, we analyze the effectiveness of texture features and incorporate them to improve our classification results. In the second part of our experiments, we address the problem of subtype classification of nephroblastoma. We want to evaluate whether there is a possibility of estimating the development of the tumor under chemotherapy. All parameters in our experiments are empirically determined.

3.1 Nephroblastoma vs. Nephroblastomatosis

We first validate the general assumption that nephroblastoma can be distinguished from their predecessors by homogeneity and size. Subsequently, we show how this distinction can be significantly improved.
For this purpose we randomly select 54 out of our 148 Wilms' tumors. We then subdivide these into 27 test and training data sets again by chance. We proceed analogously with nephroblastomatosis data sets. Since the diffuse anaplastic and necrotic subtypes are under-represented, we made sure that they occur exclusively in the test-sets. From each of our data sets we draw the middle slice of the annotated tumor region and train a random forest classifier to distinguish these two classes (nephroblastomatosis and Wilms' tumor) with 3-fold cross validation. We repeat this procedure 5 times and calculate the average accuracy at the end.

Verifying Clinical Assumptions In clinical practice, homogeneity and size of an abdominal tumor are generally used to make a distinction between nephroblastoma and nephroblastomatosis [15]. In order to validate this approach, we

Table 2. Evaluation of clinical assumptions for classification: Nephroblastoma versus Nephroblastomatosis.

	Predicted	
	Nephroblastoma	Nephroblastomatosis
Nephroblastoma	0.833 ± 0.079	0.167 ± 0.079
Nephroblastomatosis	0.185 ± 0.067	0.815 ± 0.067

calculated these two feature for all data sets and used it for classification, see Tab. 2. The average accuracy of 0.824 indicates that homogeneity and size are valuable properties to distinguish a nephroblastoma from its precursor lesion. Nevertheless, it seems not sufficient to build clinical decisions on. Thus, we add more visual texture properties to the classification procedure [8,19].

Feature Selection with Random Forests Haralick et al. [8] and Soh and Tsatsoulis [19] suggested a number of additional texture features. In a first step we calculate all of these 23 features and train a bagged random forest classifier with 300 ensemble learners. This also gives us the opportunity to evaluate the influence of each feature on the final classification. It turned out that the following nine features are decisive: size, information measure correlation 1 and 2, cluster prominence, sum entropy, dissimilarity, maximum probability, energy and autocorrelation. Surprisingly, the feature of homogeneity is not important when the above information is given. We evaluated these features as previously

Table 3. Classification result with appropriate feature selection: Nephroblastoma versus Nephroblastomatosis.

	Predicted	
	Nephroblastoma	Nephroblastomatosis
Nephroblastoma	0.926 ± 0.064	0.074 ± 0.064
Nephroblastomatosis	0.063 ± 0.027	0.937 ± 0.027

on five randomly selected data sets and 3-fold cross validation. It turns out that this additional information dramatically improves the classification performance to an accuracy of 0.932, see Tab. 3.

3.2 Subtype Determination

A Wilms' tumor consists of the tissue types stroma, epithelium, and blastema [21]. Depending on the chosen therapy strategy, the subtypes are distributed differently, see Fig. 2. In Europe, the key concept in therapy planning is a preoperative chemotherapy. This aims to shrink the tumor but also to make it

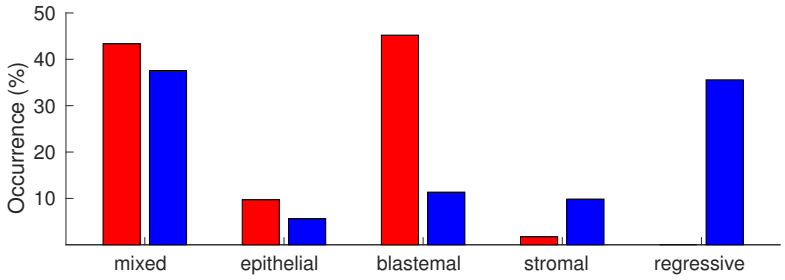

Fig. 2. Subtype distribution without (red) and with (blue) pre-operative chemotherapy.

more resistant to ruptures [6]. During this phase of therapy, various subtypes emerge, some of which differ dramatically in their prognosis. In the following we consider the standard group of intermediate risk patients. This consists of mainly regressive, epithelial dominant, stromal dominant, and mixed (none of the tissue types predominates) tumors. Since the blastemal dominant type has the worst prognosis, we also include it. Unfortunately, it is not yet possible to predict which of the subtypes develops during chemotherapy. Clinicians assume - based on subtype distributions before and after chemotherapy - that mainly blastemal tissue is destroyed during this phase of therapy, see Fig. 2. However, there is currently no possibility to determine the histological components without a biopsy, exclusively based on imaging data.

We evaluate how far we can get in subtype determination with simple but standard T_2 sequences. Since this problem is much more complex than the distinction between nephroblastoma and nephroblastomatosis, we need more data. Therefore, we select one slice from each annotated tumor from the lower third of the annotation, one from the upper third and the middle slice. In this way we generate a total of 54 images of a blastemal dominant tumor, 150 of a regressive tumor, 87 of a mixed tumor, 84 of a stromal dominant tumor and 51 of an epithelial dominant tumor.

Depending on the classification problem, we always take as many images as there are in the smaller class and divide them randomly into training and test sets. In this way we ensure that the results are not aimed at the frequency of the images but only at the discrimination. Then we calculate the visual vocabulary for each data set to generate a bag of visual words. With this information we then train a random forest with 300 ensemble learners and 3-fold cross validation. We repeat this process 5 times, analogous to the differentiation of nephroblastomatosis, including the newly generated training and test set. Here, we optimize the size of the vocabulary on the training set and select a value from the interval $[10, 100]$. We compare all selected subtypes with all others in Fig. 3. Our results are strictly above the chance level (dashed line) while average accuracy of regressive is 0.70, epithelial dominant 0.72, stromal dominant 0.66, mixed 0.67, and blastemal dominant 0.64. This indicates that we are on the right way and that it should be

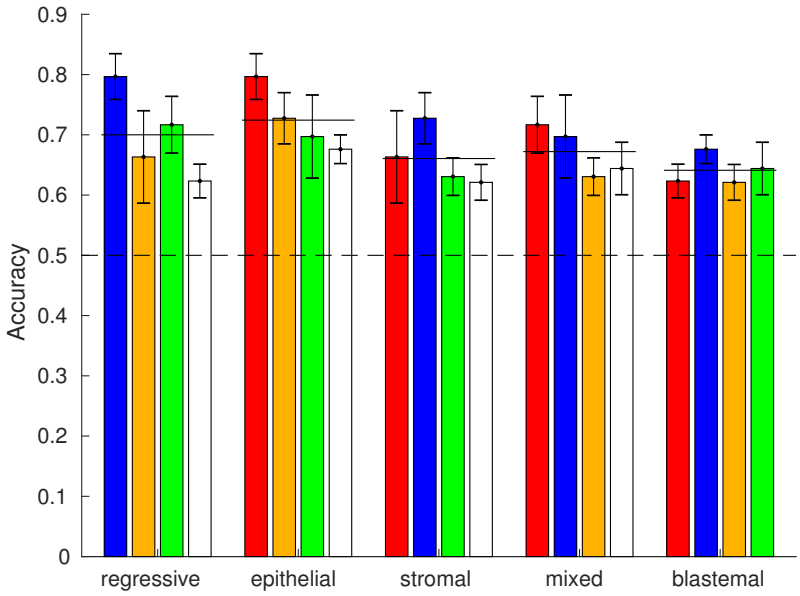

Fig. 3. Evaluation results showing mean and standard deviation of between-class classification accuracy for regressive, epithelial, stromal, mixed and blastemal subtypes. Mean performance is indicated with black lines. The dashed line marks the chance level. red: regressive, blue: epithelial, yellow: stromal, green: mixed, white: blastemal.

possible to distinguish these subtypes based on imaging data.
There are also several cases where our classification is surprisingly accurate. The accuracy of the distinction between regressive and epithelial dominant subtypes is 0.80. This leads to the following conclusions: 1. Tumors that are epithelial dominant prior to chemotherapy are less likely to regress than those that are rich in stroma or blastemal tissue. This coincides with subtype distributions before and after chemotherapy. 2. Epithelial areas can be distinguished from other types of tissue by visual features.
Furthermore, the differentiation between regressive and mixed subtypes is relatively accurate with 0.73. This allows conclusions similar to those of the epithelial type. In addition, the epithelial dominant subtype is also well distinguishable from the stromal dominant one, i.e. classification accuracy of 0.7. We also tried to use neural networks to solve our classification problem. Unfortunately it turned out that we do not have a sufficient amount of data to re-train enough layers of a pretrained network. Therefore, all our attempts with neural networks showed low performance.

We ensured that the main parameter settings of images included in our data set are as similar as possible. However, several parameters differ dramatically in

many cases. Since these cannot be compensated, the data is unfortunately not completely comparable and a considerable parameter noise is present. We firmly believe that the classification would improve significantly if this kind of noise in the data were lower. We therefore hope that in the near future a standardization of MRI sequences will be established in the medical area.

4 Conclusions

We demonstrated that the classical distinction between nephroblastomatosis and nephroblastoma is not as trivial as previously assumed. However, we were able to solve this problem by proposing further intuitive features that make the distinction much more reliable. This significantly reduces the risk of misdiagnosis and thus minimizes the medical burden on affected children.
In addition, we are the first to address the considered unsolvable problem of subtype determination prior to chemotherapy. We can show that it is basically possible to estimate this development. Even though the imaging is not standardized and therefore shows a high parameter noise, there are still visual features that allow a distinction.
Finally, we also provide the data set we use. We hope that we will be able to arouse the interest of other researchers. We hope that the estimation of the subtype in particular will be of increased interest.
In our current research, we are working on the exact visual representation of the individual classes, especially the epithelial dominant subtype. We hope that we will be able to gain more information from this in order to identify at least individual types with certainty. Most importantly, we are convinced that this research will enhance the chances of survival of the affected children. If it is possible to detect especially blastemal dominant tumors (after chemotherapy) early, the therapy can be adapted much earlier such that the recovery process of the child can be improved.

Acknowledgements. J. Weickert has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 741215, ERC Advanced Grant INCOVID).

References

1. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: Speeded up robust features. In: Proc. 2006 European Conference on Computer Vision. pp. 404–417. Lecture Notes in Computer Science, Springer (2006)
2. Beckwith, J.B., Kiviat, N.B., Bonadio, J.F.: Nephrogenic rests, nephroblastomatosis, and the pathogenesis of Wilms' tumor. Pediatric Pathology **10**(1-2), 1–36 (1990)
3. Breiman, L.: Random Forests. Machine Learning **45**(1), 5–32 (2001)

4. Cox, S.G., Kilborn, T., Pillay, K., Davidson, A., Millar, A.J.: Magnetic resonance imaging versus histopathology in Wilms tumor and nephroblastomatosis: 3 examples of noncorrelation. Journal of Pediatric Hematology/Oncology **36**(2), e81–e84 (2014)
5. Davidoff, A.M.: Wilms' tumor. Current Opinion in Pediatrics **21**(3), 357–364 (2009)
6. Graf, N., Tournade, M.F., de Kraker, J.: The role of preoperative chemotherapy in the management of Wilms' tumor: The SIOP studies. Urologic Clinics of North America **27**(3), 443–454 (2000)
7. Gylys-Morin, V., Hoffer, F., Kozakewich, H., Shamberger, R.: Wilms' tumor and nephroblastomatosis: imaging characteristics at gadolinium-enhanced MR imaging. Radiology **188**(2), 517–521 (1993)
8. Haralick, R.M., Shanmugam, K., et al.: Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics (6), 610–621 (1973)
9. Hötker, A.M., Mazaheri, Y., Lollert, A., Müller, S., Furtwängler, R., Schenk, J.P., Düber, C., Weyer-Elberich, V., Akin, O., Graf, N., Staatz, G.: Diffusion-weighted MRI in the assessment of nephroblastoma: Results of a multi-center trial. submitted (2018)
10. Kaste, S.C., Dome, J.S., Babyn, P.S., Graf, N.M., Grundy, P., Godzinski, J., Levitt, G.A., Jenkinson, H.: Wilms' tumour: prognostic factors, staging, therapy and late effects. Pediatric Radiology **38**(1), 2–17 (2008)
11. Kim, S., Chung, D.H.: Pediatric solid malignancies: neuroblastoma and Wilms' tumor. Surgical Clinics of North America **86**(2), 469–487 (2006)
12. Lloyd, S.: Least squares quantization in PCM. IEEE Transactions on Information Theory **28**(2), 129–137 (Mar 1982)
13. Lonergan, G.J., Martinez-Leon, M.I., Agrons, G.A., Montemarano, H., Suarez, E.S.: Nephrogenic rests, nephroblastomatosis, and associated lesions of the kidney. Radiographics **18**(4), 947–968 (1998)
14. Müller, S., Ochs, P., Weickert, J., Graf, N.: Robust interactive multi-label segmentation with an advanced edge detector. In: Andres, B., Rosenhahn, B. (eds.) Pattern Recognition, Lecture Notes in Computer Science, vol. 9796, pp. 117–128. Springer, Cham, Switzerland (2016)
15. Owens, C.M., Brisse, H.J., Olsen, Ø.E., Begent, J., Smets, A.M.: Bilateral disease and new trends in Wilms tumour. Pediatric Radiology **38**(1), 30–39 (2008)
16. Pastore, G., Znaor, A., Spreafico, F., Graf, N., Pritchard-Jones, K., Steliarova-Foucher, E.: Malignant renal tumours incidence and survival in European children (1978–1997): Report from the Automated Childhood Cancer Information System project. European Journal of Cancer **42**(13), 2103–2114 (2006)
17. Reinhard, H., Schmidt, A., Furtwängler, R., Leuschner, I., Rübe, C., Von Schweinitz, D., Zoubek, A., Niggli, F., Graf, N.: Outcome of relapses of nephroblastoma in patients registered in the SIOP/GPOH trials and studies. Oncology Reports **20**(2), 463–467 (2008)
18. Rohrschneider, W.K., Weirich, A., Rieden, K., Darge, K., Tröger, J., Graf, N.: US, CT and MR imaging characteristics of nephroblastomatosis. Pediatric Radiology **28**(6), 435–443 (1998)
19. Soh, L.K., Tsatsoulis, C.: Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. CSE Journal Articles p. 47 (1999)
20. Soomro, M.H., Giunta, G., Laghi, A., Caruso, D., Ciolina, M., De Marchis, C., Conforto, S., Schmid, M.: Haralick's texture analysis applied to colorectal T2-weighted MRI: A preliminary study of significance for cancer evolution. In: Proc. 13th International Conference on Biomedical Engineering. pp. 16–19. IEEE (2017)

21. Vujanić, G.M., Sandstedt, B.: The pathology of Wilms' tumour (nephroblastoma): the International Society of Paediatric Oncology approach. Journal of Clinical Pathology **63**(2), 102–109 (2010)
22. Vujanić, G.M., Sandstedt, B., Harms, D., Kelsey, A., Leuschner, I., de Kraker, J., Committee, S.N.S.: Revised International Society of Paediatric Oncology (SIOP) working classification of renal tumors of childhood. Medical and Pediatric Oncology **38**(2), 79–82 (2002)
23. Wibmer, A., Hricak, H., Gondo, T., Matsumoto, K., Veeraraghavan, H., Fehr, D., Zheng, J., Goldman, D., Moskowitz, C., Fine, S.W., et al.: Haralick texture analysis of prostate MRI: utility for differentiating non-cancerous prostate from prostate cancer and differentiating prostate cancers with different gleason scores. European Radiology **25**(10), 2840–2850 (2015)
24. Zayed, N., Elnemr, H.A.: Statistical analysis of Haralick texture features to discriminate lung abnormalities. Journal of Biomedical Imaging **2015**, 12 (2015)