Textural and deep learning methods in recognition of renal cancer types based on CT images

Aleksandra Maria Osowska-Kurczab*, Tomasz Markiewicz*†, Miroslaw Dziekiewicz‡ and Malgorzata Lorent†
* Faculty of Electrical Engineering
Warsaw University of Technology, Poland
† Department of Pathology
Military Institute of Medicine, Warsaw, Poland
‡ Department of Vascular and Endovascular Surgery
Military Institute of Medicine, Warsaw, Poland
Email: olaosows@gmail.com, markiewj@wp.pl, dziekiewicz@wp.pl, mlorent@wim.mil.pl

***Abstract*—Recent advancements in deep learning have opened new prospects in many areas of research. Especially interesting field is biomedical image analysis, where plenty of problems wait for efficient solution. The aim of this work is to develop new approaches to recognition of different types of renal cancer on the basis of Computed Tomography (CT) imaging. Two different directions will be investigated. One uses the texture descriptors of the images to define the diagnostic features. They are next combined with support vector machine responsible for final recognition and classification. The second applies deep learning approach using different configurations of Convolutional Neural Networks. The experimental research for both textural and deep learning approaches was conducted on real world dataset of CT scans consisting of eight types of renal cell carcinomas. The proposed structures of predictive system were able to achieve the level of accuracy around 90% for complex and unbalanced datasets.**

***Index Terms*—Computer Vision, Deep Learning, Convolutional Neural Networks, textural features, Support Vector Machine, medical imaging, renal cancer**

I. Introduction

The ratio of deaths caused by the wide range of neoplastic diseases was steadily increasing in the last decade. Cancer nowadays is called a disease of affluence, because of high correlation between becoming ill and unhealthy lifestyle, environmental pollution and stress factors. Meanwhile, many regions of the world are currently facing problems with shortage in medical professionals. These issues result in prolonging queues to specialist's appointment. However, passing time is the main enemy for oncological diseases, because of the fact that the ability of curing the patient is strictly associated with his initial condition at the first visit to the cancer specialist.

For this reason, the research in supporting the medical diagnosis by automated systems is very important. Having in mind major advances in computational technology and development of more and more sophisticated methods in machine learning, the possibility of creating systems, used in early diagnostics, is rising. The scientific society should work hand in hand with medical staff to devise systems that are as helpful as possible.

This work was financially supported by National Science Centre, Poland (grant no 2016/23/B/ST6/00621).

This article will address the problem of creating a system of automatic analysis of Computed Tomography (CT) scans of patients with Renal Cell Carcinomas (RCC). The main task is classifying the Regions of Interest (ROI) to the specified RCC type. The task will include the process of creating data set, extracting ROI and methods of data sets augmentation. Key aspect considered in the research is comparison of two groups of methods in creating features for pattern recognition algorithms. The first one applies texture analysis methods, well-established in the literature. It is based on some structural, statistical or transformation information derived from raw grey-scale images. The second one relies on Transfer Learning (TL) technique applied to popular pretrained networks. Despite the fact, that the collected RCC dataset is rather small and unbalanced (because of scarce representation of patients with rare variants of RCC), aforementioned technique enables to create models with comparable or promising results to those based on textural features.

II. Previous work

This section will briefly present few exemplary applications of both textural and neural-based approaches. Quite popular is the problem of differentiation between benign and malignant variants of kidney tumours. For this binary classification task, four numerical features: contrast, correlation, energy and homogeneity based on Gray-Level Co-occurrence Matrix (GLCM) and passing them to classical classifiers like SVM or KNN turns out to be sufficient [1]. Numerical features of the images can be also computed using histograms of curvature-related features, histogram of oriented gradients or raw intensities of pixels. Quite popular in literature is the Computed Tomography Texture Analysis (CTTA) approach [2], [3] which applies the computation of simple features of raw image pixel values in order to obtain biomarkers for tumour differentiation. Such features include average intensity, entropy, skewness, kurtosis, standard deviation, histograms of pixel intensities or frequency analysis results.

Combining Gabor filtering and ensemble classifiers with majority voting can improve results of textural analysis of RCC as shown in [4]. With extended domain knowledge about

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

RCC types similarity, cascade of binary classifiers [5] can be created instead of single multi-class classifier. This technique is intended to be mimicking the steps undertaken by medical doctors.

When neural based methods are considered, Deep Belief Networks and Autoencoders [6] are used for some tasks in the area of biomedical image analysis. Tedious stage of features generation can be replaced with Convolutional Neural Networks (CNN). Interesting results were obtained by applying 3D-CNN models to classification of lung tumours [7], [8]. 3D-CNN version of popular pretrained 2D images models like AlexNet or GoogleNet were presented. Authors of this research indicated possibility of improving the overall performance with hybrid models. Convolutional part of 3D-CNN model, which is responsible for features generation, is extracted and then combined with classical classifiers like SVM, XGBoost or Random Forest. However, it is worth noting, that learning deep network from scratch is a process requiring large number of learning examples.

Equally important to the classification task is also the segmentation problem. Main purpose of the segmentation process is extracting ROI from full frame of the scan. This ROI is usually a homogeneous region with similar properties e.g. a cancer region or a normal organ. Segmentation algorithms can be divided into two subgroups: non-trainable (e.g. thresholding, watershed transformation or edge detection) and trainable (U-net [9], CNN [8] and fuzzy c-means clustering). Segmented region can be used to obtain information about size and volume of the tumour. These are important parameters while planning pharmacological therapy.

Up to now, no papers have been devoted to recognition of different types of renal carcinoma. It is very difficult problem due to large variety of the same type of changes in lesion regions of different patients, as well as very small areas representing such lesions.

III. Texture Analysis Methods

Texture analysis enables the characterisation of lesion regions with the numerical values based on raw intensities of the grayscale image. It looks for such properties as roughness, irregularity or smoothness. Results of texture analysis, provide feature vectors for classification, segmentation or detection task. In this research four carefully selected, distinct algorithms were tested and compared to deep learning approach.

A. GLCM features

Gray-Level Co-Occurrence Matrix is one of the most popular algorithms of texture analysis [10]. It still finds application in analysing biomedical objects e.g. biological cells, x-ray or CT images. Key aspect of the GLCM algorithm is representing the image from the perspective of relation and distribution of gray-levels of pixels. Numerical descriptors are computed on the basis of co-occurence matrix $h_{d\theta}$, where each position (i, j) in this matrix $h_{d\theta}(i, j)$ represents the number of co-occurences of intensity values i and j of the image, assuming some offset function. In this research multiple symmetrical offset function were tested with distances ranging from $d \in \{1, 2, 3\}$ and angle $\theta \in \{0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°\}$. The list of originally proposed descriptors contained fourteen different parameters, e.g. energy, contrast, homogeneity, entropy, variance and correlation coefficient. This list can be enriched by statistical measures such as median, kurtosis or skewness. Final feature vector included fourteen well selected predictive features. After determining the list of descriptors, the feature vector should be normalized. In this research, standardization (z-score) was used, which is normalization using mean and standard deviation.

B. Fractal features

Fractal methods [11] are commonly applied to objects with high local complexity. Fractals themselves are described with fractional Brownian Motion (fBM) model or extended self-similar (ESS) model, which is a generalization of fBM. However, structures with strong orientation dependences might not be modelled well with fractal features. Fortunately, RCC images don't show strong orientational dependences. The SFTA algorithm proposed in [12] consist of two separate stages: decomposition of input image into a set of binary images with Two-Threshold Binary Decomposition and computing fractal dimensions of the regions borders. Length of the initial feature vector generated by SFTA algorithm will be denoted with n and is subject to changes. Final feature vector included 33 features.

C. Unser features

Unser features [13], [14] were often used in textural segmentation task. Linear filter is applied to an image, forming the sum or difference of shifted image parts. On the basis of this matrix, the descriptors such as variance, energy, correlation, contrast and many other popular measures, are computed. In this research 8 best features were selected. Two hyperparameters should be chosen while using Unser method - shift length and filter size.

D. Gabor filtering

This type of filtering is widely used in initial transformation of textural images. Gabor filter is a specific type of band-pass filter, which is susceptible to orientation. Therefore, this filtration method is quite popular in such tasks as textural analysis or edge detection. One of the most important properties of this method is its ability to generate informative description for various shapes, sizes and smoothnesses of the image.

In this research, multiple types of descriptor vectors were defined on the bases of Gabor filtering results:

- basic vector consisting of energy, mean amplitude and entropy,
- previous vector exteded by average intensity, skewness, kurtosis and standard deviation,
- local binary patterns (LBP) features [15] combined with extended vector,

- histograms of oriented gradients (HOG) features [16] combined with an extended vector.

In our solution the descriptors were based on real part of the filtered image (found as the best in experiments). Depending on the initial type of feature generation strategy, final vector of descriptors included from 120 to 280 features.

IV. Deep Learning Models

Convolutional Neural Networks are deep multilayer structures [17]–[19], which are tackling two tasks at the same time: generation of features for modelled problem and ultimate prediction (e.g. regression or classification). Fact that tedious process of feature analysis is done automatically has significantly contributed to general popularization of CNNs in almost any pattern recognition tasks. The only bottleneck in developing CNN models is large amount of data needed to train such network.

Transfer learning is a method of training deep CNNs that is solving this main obstacle. Instead of training all weights of the neural connections with randomly initialized values, a model trained on another task B is adapted to fit to the new task. Weight adaptation speed is much higher than learning from scratch. Additional advantage of transfer learning is the fact, that if the task A and B are sufficiently similar, amount of data needed to retrain the network is significantly smaller than to train from the beginning.

Adaptation of model to the particular task can be performed in multiple ways. Crucial part of every transfer learning is to substitute the final predictor part of pretrained network (layers such as fully connected, softmax, classification output). This part has to be fit to a new task (e.g. the number of classes in classification task). Feature extraction part can be reused either without any adjustments (layers are freezed) or can be altered through training (initial weights are not random).

Transfer learning in our task has been applied to well-known CNN models, such as AlexNet [20], ResNet [21], Inception [22], [23] and Inception-ResNet [24]. All of these networks were competing in ILSVRC Challenge and have obtained state of the art results. They were trained on ImageNet database containing of more than 1 million RGB images divided into 1000 categories such as *'clock tower'*, *'vitamin pill'* and *'electric guitar'*. The architectures evolved and improved throughout the years which resulted in significant progress in image pattern recognition. The main features of these networks are listed below.

A. AlexNet

- nonlinear activation function tanh superseded by ReLu to speed up the computation
- dropout as regularization tool to avoid overfitting
- overlapping pooling for network size reduction

B. ResNet

- vanishing gradients (training convergence problem in deeper networks) solved with residual connections
- batch normalisation and global average pooling used in layer structures

C. Inception

- vanishing gradients solved with additional classifiers added in inner layers
- inception cell implementation as stacks of convolutions 1x1, 3x3, 5x5 organised in parallel branches

D. InceptionResNet

- hybridization of inception cells and residual connections concepts to improve overall performance

V. Data Sets and Methodology

Aforementioned methods of features generation will be applied in the task of recognising neoplastic renal lesion types. These lesions include RCC, renal cysts and tumours of different origin, located in kidney. RCC belongs to the atypical cancers. For example, in the UK it is the 7th most common one [25]. However, its proportion to other cancers is still growing, due to more accurate diagnostics in the last decade. Vast majority of the RCC cases are recognised accidentally during diagnosing other diseases. It is due to the fact that early stages of this disease rarely cause any symptoms, hence it is much less specific. These facts show that automation of CT scans analysis might significantly reduce the number of cases diagnosed in advanced stages.

The most common type of renal lesions is Clear Cell Renal Cell Carcinoma (CC-RCC) [27], which is responsible for 65-70% of all renal cancer cases. CC-RCC has a high metastases incidence rate to such organs as lung (75%), liver (40%), bone (40%), soft tissue (34%) and pleura (31%). This type of RCC has the worst prognosis among all RCC with respect to such diagnostic variables as nucleolar prominence (Grade 1, 2 and 3) or the presence of pronounced nuclear pleomorphism (Grade 4). In CC-RCC the 10-years survival rate varies from 96%, 90%, 50% to 20% from grade 1 to 4 respectively. The next frequently identified RCC is papillary renal cell carcinoma (PRCC) that accounts for approximately 15% of RCC cases. It has more favourable survival outcome than CC-RCC with the 10-years survival changing from 100%, 94%, 78% to 37% for grade 1 to 4 respectively. The chromophobe renal cell carcinoma (ChRCC) constitutes 5-7% of all RCC. It has a good prognosis with 5-year survival rate of 78-100%. This type of RCC cannot be graded yet.

In contrast to RCC, two benign renal tumours would be taken into account in this study. The first one is oncocytoma [28] composed predominantly of large eosinophilic cells packed with mitochondria. This benign epithelial tumour located in kidney is responsible for 5-9% of all renal neoplasm cases. When no other co-morbid RCC disease is present, a 100% survival rate can be observed. The second benign renal neoplasm considered is angiomyolipoma with approximately 1% of surgically removed renal tumours. It is a benign mesenchymal tumour composed of varying proportion of adipose tissue, spindle cells, muscle cells and abnormal blood vessels. Small minority of cases are associated with complications and mortality. The last type of renal lesion taken into account in

this study is renal cyst, which is not a malignant itself, but should be differentiated from the renal tumours.

A. Data source and renal cancer types

In this research a dataset of scans belonging to 143 patients with 8, the most popular types of renal lesions was gathered in cooperation with Military Institute of Medicine (Warsaw, Poland). Table I shows the names of renal lesion types, their shortened notations and population with respect to number of patients and number of extracted frames. Examples of frames containing kidney and cancer outlines representing each class are depicted in Fig. 1.

It is evident that available data set is strongly unbalanced. Number of frames in the most numerous class (J) is over 7 times higher than in the least populated class (A). It is the result of differences in disease categories incidence rates, which is the main obstacle in gathering balanced dataset. Other issues that should be mentioned are the size and information density carried in single grayscale frames of CT. Average size of tumour bounding box, regardless of RCC type, is only 25x25px. Therefore, information conveyed by small images of low contrast is rather poor. A single frame might be even ambiguous for a professional oncologist. However, assembled examples in the classes are believed to be representative and of good quality, though they are coming from different CT scanners.

TABLE I
QUANTITY OF RENAL LESIONS IN THE DATA SET

Full RCC name	Abb. name	Number of patients	Number of scan frames
Angiomyolipoma	A	8	97
Chromophobe Renal Cell Carcinoma	C	20	253
Clear Cell Renal Cell Carcinoma	J	40	692
Multilocular Cystic Renal Cell Carcinoma	M	10	164
Oncocytoma	O	14	108
Papillary Renal Cell Carcinoma	P	26	236
Urothelial Carcinoma	R	11	292
Renal Cystis	T	14	460

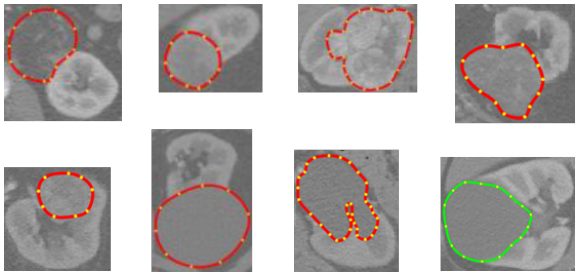


Fig. 1. Example frames containing kidney and cancer contour for each of the classes (from the upper left corner: A, C, J, M, and below: O, P, R, T).

Outlines of the tumour regions were prepared by a surgeon with oncological experience. Histopathology descriptions were also available for inspection. Results of exploratory data analysis have shown that 65% of patients are male with average age 64 years. Around 60% of the scans come from the year 2015 or later.

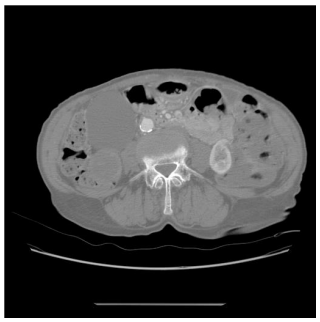
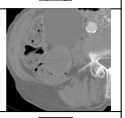
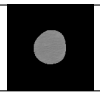
B. Training data preparation

Preparation of training and testing data was composed of three main steps: normalisation, ROI extraction and data sets augmentation. The first step uses dicom file parameters such as window width, window center and rescale intercept to adjust the values to a common scale uint8.

Second step intends to reduce the full frame of the scan to more distinct representations, called ROI. In this research, we assume that segmentation task is already done, and position of the tumour is known from contours prepared by medical specialist. However, the ROI can be extracted from full scan frame in many ways. In table II five different types of extraction methods are illustrated (denoted by D1, D2 up to D5).

- D1 - Bounding box of RCC contour without surrounding tissues,
- D2 - Bounding box of RCC contour with surrounding tissues,
- D3 - 150x150px region where centre is in bounding box centroid, with surrounding tissues,
- D4 - 50x50mm region where centre is in centroid of bounding box, without surrounding tissues,
- D5 - 100x100mm region where centre is in centroid of bounding box, without surrounding tissues.

TABLE II
ILLUSTRATION OF 5 TYPES OF ROI

Original frame	ROI type	Extracted ROI
	D1	
	D2	
	D3	
	D4	
	D5	

The influence of the chosen extraction method on the quality of the proposed classification system will be studied. It is aimed to reduce the impact of class imbalance on quality of classifier performance. The augmentation process takes into consideration the fact, that images are small and shouldn't be altered too much to prevent major distortions. Two basic transformations were used: rotation and cropping. Finally, every class was augmented to have representation of at least 500 frames.

C. Experimental settings

10-fold cross validation technique was used to analyse ultimate performance of trained models. To compare different methods the classical quality measures defined on the basis of confusion matrix, such as accuracy, weighted precision, recall and F1-score [29] were used.

Two types of experiments have been performed. The first one was based on texture analysis in feature generation and application of Support Vector Machine (SVM) [26] as classifier. Parameters of SVM were chosen as follows: Gaussian kernel with $C = 1000$ and $\gamma = 1$).

In the second case the generation of features and classification phases were combined in deep learning structures, representing various implementations. Transfer learning was applied in this approach. The preliminary research was needed to find relationship between cut-off layer and output performance. For all networks, similar pattern was observed: the earlier freeze of the weights, the better results of recognition. Final models were trained in ensemble manner with majority voting. Due to limitation of computation resources, the pretrained models were divided into two subgroups. The first one consisted of AlexNet. Ten independent networks were separately trained and freezing was not applied to any parameters of the layers. The second group was created by all other models (ResNet, Inception and InceptionResNet). Ensemble model was formed of 5 individual predictors and freezing was applied to early convolutional layers.

Independence of ensemble members was provided thanks to 2 different mechanisms:

- Random choice of proportion of training and validation sets in learning folds in 10-fold cross validation scheme for each ensemble member. This proportion has been changing from 3:1 to 9:1.
- Different sizes of fully connected layer for each of ensemble members.

Training was performed using Adam solver with small 10 element batches and learning rate scheduler with initial value 1e-4.

VI. Results

A. Texture analysis results

Comparison of results of all texture analysis experiments is presented on bar plot in the Fig. 2. F1-scores for every feature method were presented with respect to five types of data set (D1-D5) and four methods of feature generation (GLCM, fractal, Unser, Gabor).

More detailed information about top 3 results can be also found in table III. The best results were obtained for fractal features and D5 type of data set. Weighted F1-score reached the value close to 90%. Fractal method appeared to outperform almost all other methods, regardless of the data set. ROI extraction method named D5 has resulted into the best quality measures for almost all texture analysis methods.

One of the most import conclusion from results presented in Fig. 2 is the big influence of ROI generation techniques

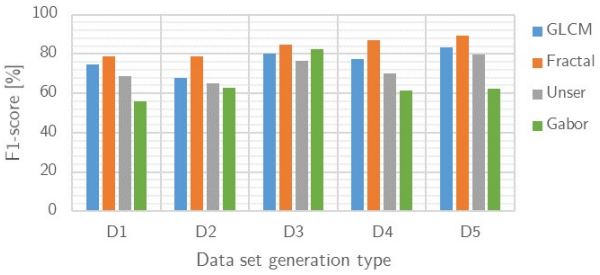


Fig. 2. Results of texture analysis experiments with respect to 4 different types of texture description.

TABLE III
TOP 3 RESULTS OF TEXTURAL ANALYSIS EXPERIMENTS (IN %)

No	Method	Data set	Acc	Prec	Rec	F1
1	Fractal [a]	D5	89.0	89.7	89.0	89.1
2	Fractal [b]	D5	88.5	88.8	88.5	88.5
3	GLCM [c]	D5	80.9	81.5	81.1	81.0

[a] $n = 7$
[b] $n = 5$
[c] $d = \{1, 2, 3\}$, full θ range, averaging tangent directions

on overall classification performance. It is evident that not only feature selection, but also way of extracting ROI region is crucial in creating efficient classification system of renal images. As fractal features are concerned, minimal F1-score result equals 76.2% for D2 data set and maximum for D5 - 89.1% (13 percentage points difference).

10 fold cross validation gave ultimate F1-score results on the level of 89.1% $\pm$ 1.2%. Throughout cross validation process, the maximal obtained F1 value was 90.6%, minimal 86.4% with median equal 89.1%. The aggregated confusion matrix is presented in Fig. 3. Most mistakes are related to recognition of majority class J. Perhaps, it was caused by very large number of different patients from which the images were collected and significant differences among images for each of them. According to medical experts RCC types are usually misclassified in two separate groups: either between classes C-J-P or P-R-T. This patterns can be also observed in Fig. 3.

To inspect the results more precisely, bar plot with F1-scoring versus class is presented in Fig. 4. Classes A, O and M are the ones with results above the average, whereas classes J, P, R and T are the most problematic in recognition.

Fractal model outperforms other textural algorithms of feature extraction. Presumably, it might be due to the fact that SFTA algorithm is intended to be used in describing textures with strong local features, which is the case of RCC renal types.

B. Deep learning results

Six different arrangements of transfer learning have been investigated. The pretrained networks used in experiments were as follow: AlexNet_1 (single AlexNet), AlexNet_10 (ensemble of 10 separately trained AlexNets), ResNet-18_5 (ensemble of 5 separately trained ResNet-18), ResNet-50_5, Inception-v3_5

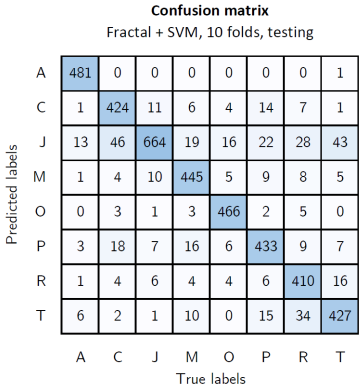


Fig. 3. Confusion matrix for fractal experiment.

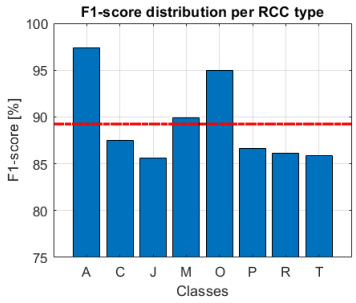


Fig. 4. Distribution of F1-scores for fractal method.

and Inception-ResNet-v2_5 (same arrangement as for ResNet-18).

The results of recognition of eight types of RCC in the form of F1 measure are presented in a bar plot in Fig. 5. They correspond to the testing results obtained in 10-fold cross validation mode by the ensembles integrated with majority voting rule. Detailed information about top 3 results can be also found in table IV.

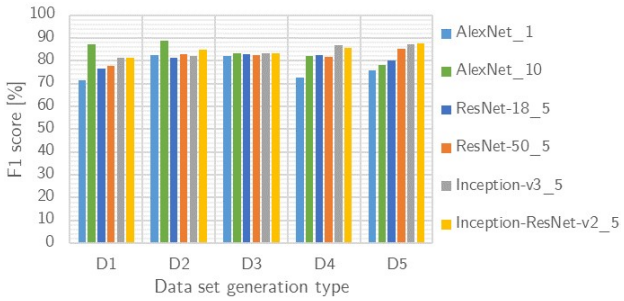


Fig. 5. Results of deep learning experiments with respect to 5 ROI extraction types and 6 types of networks.

The best results were obtained in ensemble composed of 10 AlexNet predictors on D2 data set. F1-score has reached the level of $87.1\% \pm 1.6\%$. The maximal difference of results in AlexNet ensemble for various data sets was 8 percentage points (between D2 and D1). 10-fold cross validation varied from 83.7 % to 89.2% with median of 87.3%. Mostly misclassified examples belonged to classes J and T.

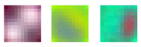
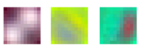
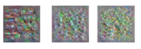
TABLE IV
TOP 3 RESULTS OF DEEP LEARNING EXPERIMENTS (IN %)

No	Method	Data set	Acc	Prec	Rec	F1
1	AlexNet_10	D2	87.1	87.5	87.1	87.1
2	Inception-ResNet-v2_5	D5	86.4	86.7	86.4	86.5
3	Inception-v3_5	D4	85.7	86.0	85.3	85.7

As shown in Fig. 5, ensemble technique has improved the classification results by almost 7 percentage points comparing to the best individual classifier. Diversity of predictor's results within an ensemble provide the space for compensation of some errors. The only drawback of such solution is the increase of computation time.

Though, the RCC data set differs in quality, size and information density from ImageNet database, transfer learning achieved comparable results with baseline textural methods. The way the features are generated in pretraining stage are already applied in retraining process (transfer learning) on the RCC data set. It is interesting to observe the changes of the weight values of filters in different layers of CNN in the process of retraining the structure starting from the pretrained form. It is presented in table V for three image filters taken from the first, second and fifth convolutional layer. Upper row depicts a visual representation of weights in 3 chosen convolutional filters of different layers. The bottom row represents their values after retraining. The deeper in the network layers the more adjustments are made to the initial weights in order to adapt the network to the new task.

TABLE V
VISUALIZATION OF CHANGES IN WEIGHTS VALUES

	conv1	conv2	conv5
pretrained			
retrained			

C. Discussion

Both investigated methods based on texture analysis and deep learning have shown high efficiency in recognition of renal lesion types. Quality measures (accuracy, precision, recall and F1) have reached high scores, which is good prognostics for application in medical diagnostic support. Textural methods have been found vulnerable to many parameters of the system - ROI generation method, selection of hyperparameters values and diagnostic features. These disadvantages are partly eliminated in Deep Learning approach. However, ROI generation method and choice of architectural parameters are equally important.

Summarizing, we may state that both methods deliver results of comparable quality. The decision which method should be chosen in particular application may depend on

user expectations and limitations. The inference time of texture based models is relatively shorter in contrast to ensemble of CNN, which is more computational demanding. On the other hand, Deep Learning methods do not require sophisticated expert knowledge needed in creation of diagnostic features. Thus, it is more convenient in practical applications.

It should be noticed that differentiation between RCC types is in fact impossible for a person without specialised medical education. Spotting small highlights and subtle textural nuances among different types of RCC turns out to be a very complex and prone to errors problem. Therefore, the development of computerised recognition system is very useful to support the medical staff in taking proper diagnostic decision.

VII. Summary and future work

In this article we present a complete method of creating an automatic system of medical image analysis in the task of renal tumours recognition. Two approaches to this problem were presented. Both textural and deep learning methods developed in the paper have provided satisfactory results, with accuracy reaching 90%.

The results of this study might find application in supporting the medical diagnosis in hospital practice. Implementation of such system in automatic medical images analysis could lead to significant acceleration of the diagnosis process and to better prioritization of patients waiting for consultation with medical specialists.

The important task in the future research, is to create more balanced data sets and increase the total number of CT scans. Scarce representation of some classes has a negative impact on model outcome. Additional problems that have been mentioned in this study are related to small size of lesion region, wide inter-patient diversity and limited information density of images. Hence proper segmentation of images, careful selection of features and choice of classification system are of great importance.

The size of the tumours is right now a bottleneck in improving deep learning models. Resizing small images to the required size of input layer might distort initial information hidden in the CT scan fragment. Finding an effective method of resizing and improving resolution is an inevitable step in further development of deep learning models. Techniques of super resolution, deblurring or adding deep image prior would be investigated.

Providing that a bigger dataset would be collected, 3D CNN model will be developed. Perhaps, information about the sequence of frames can refine the overall performance. It is worth noting that medical experts derive final diagnosis from series of images, not from single static frame.

In the nearest future, more attention will be drawn to method of ROI generation, data set augmentation algorithm and feature selection. Major performance improvement is expected when multi-feature ensemble model is applied. Combining various methods of feature generation with few classification algorithms might help to minimize the ratio of misclassified examples.

Finally, this research is intended to provide helpful information to the medical doctors about the patient condition rather than replace classical (expert) medical diagnosis process. In the nearest future, explainability of the classification models will be explored with techniques such as Grad-CAM. Supporting medical diagnosis by automated system might have significant influence on diagnosis durance - especially when the model is self explaining its decisions with insightful information for the medical specialist.

References

[1] M. Sbert, "Kidney Tumor Segmentation and Classification on Abdominal CT Scans" in International Journal of Computer Applications (0975 – 8887), vol. 164, no. 9, 2017.

[2] B. Kocak et al., "Textural differences between renal cell carcinoma subtypes: Machine learning-based quantitative computed tomography texture analysis with independent external validation" in European Journal of Radiology, vol. 107, pp. 149-157, 2018.

[3] S. P. Raman, Y. Chen, J. L. Schroeder, P. Huang, E. K. Fishman, "CT texture analysis of renal masses: pilot study using random forest classification for prediction of pathology" in Academic Radiology, vol. 12, pp. 1587-96, 2014.

[4] L. Mredhula, "Detection and classification of tumors in CT images" in Indian Journal of Computer Science and Engineering (IJCSE), vol. 6, no. 2, 2015.

[5] M. G. Linguraru, S. Wang, F. Shah, R. Gautam, J. Peterson, W. M. Linehan, R. M. Summers, "Automated noninvasive classification of renal cancer on multiphase CT" in Medical physics, vol. 38, no. 10, pp. 5738–5746, 2011.

[6] W. Sun, B. Zheng, W. Qian, "Computer aided lung cancer diagnosis with deep learning algorithms" in Proceedings of the International Society for Optics and Photonics Conference, 2016.

[7] H. Polat, D. M. Homay, "Classification of Pulmonary CT Images by Using Hybrid 3D-Deep Convolutional Neural Network Architecture" in Applied Sciences, vol. 9, no. 5, 2019.

[8] W. Alakwaa, M. Nassef, A. Badr, "Lung cancer Detection and Classification with 3D Convolutional Neural Network (3D-CNN)" in International Journal of Advanced Computer Science and Applications, IJACSA 2017, vol. 8, no. 8, 2017.

[9] O. Ronneberger, P. Fischer, T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation" in Medical Image Computing and Computer-Assisted Intervention, MICCAI 2015, vol. 9351, 2015.

[10] R. M. Haralick, K. Shanmugam and I. Dinstein, "Textural Features for Image Classification" in IEEE Transactions on Systems, Man and Cybernetics, vol. SMC-3, no.6, 1973.

[11] P. Shanmugavadivu and V. Sivakumar, "Fractal Dimension Based Texture Analysis of Digital Images" in Procedia Engineering, vol. 38, pp. 2981-2986, 2012.

[12] A. F. Costa, G. Humpire-Mamani and A. J. M. Traina, "An Efficient Algorithm for Fractal Analysis of Textures," 2012 25th SIBGRAPI Conference on Graphics, Patterns and Images, Ouro Preto, 2012, pp. 39-46.

[13] M. Unser, "Sum and difference histograms for texture classification", IEEE Transactions on Pattern Analysis and Machine Intelligence, p.118-125, 1986.

[14] M. Unser, "Local Linear Transforms for Texture Analysis" in Proceedings of the 7th IEEE International Conference on Pattern Recognition, ICPR 1984, vol. II, pp. 1206-1208, 1984.

[15] T. Ojala, M. Pietikäinen, and D. Harwood, "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions" in Proceedings of the 12th IAPR International Conference on Pattern Recognition, ICPR 1994, vol. 1, pp. 582-585, 1994.

[16] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection" in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, 1, vol. 1, pp. 886-893, 2005.

[17] Y. Bengio, "Learning Deep Architectures for AI" in Foundations and Trends in Machine Learning, vol. 2, no. 1, pp. 1–127, 2009.

[18] Y. Bengio, Y. LeCun, G. Hinton, "Deep Learning" in Nature, vol. 521, pp. 436–444, 2015.

[19] I. Goodfellow, Y. Bengio, A. Courville, "Deep Learning", MIT Press, 2016.

[20] A. Krizhevsky, I. Sutskever and G. Hinton, "Image net classification with deep convolutional neural networks" in Advances in Neural Information Processing Systems, vol. 25, pp. 1-9, 2012.
[21] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition" in Proceeding of the 29th IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2016, pp. 770-778, 2016.
[22] C. Szegedy et al., "Going deeper with convolutions" in Proceedings of the 28th IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2015, pp. 1-9, 2015.
[23] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision" in Proceedings of the 29th IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2016, pp. 2818-2826, 2016.
[24] C. Szegedy, S. Ioffe, V. Vanhoucke and A. Alemi, "Inception-v4, Inception-ResNet and the impact of residual connections on learning" in Proceedings of 31st Association for the Advancement of Artificial Intelligence on Artificial Intelligence, AAAI 2016, pp. 1-12, 2016.
[25] "Kidney cancer statistics", Cancer Research UK, 2020. [Online]. Available: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/kidney-cancer. [Accessed: 31-Mar-2020].
[26] C. Cortes and V. Vapnik, "Support-vector network" in Machine Learning, 3, vol. 20, pp. 273-297, 1995.
[27] H. Moch, A. L. Cubilla, P. A. Humphrey, V. E. Reuter, T. M. Ulbright, "The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs-Part A: Renal, Penile, and Testicular Tumours" in European Urology, Volume 70, Issue 1, 93 - 105, 2016.
[28] T. Gudbjartsson et al., "Renal oncocytoma: a clinicopathological analysis of 45 consecutive cases" in BJU International, 96: 1275-1279, 2005.
[29] P. N. Tan, M. Steinbach, V. Kumar, "Introduction to Data Mining", Pearson Education, Boston, 2006.