2023-10

Kidney Tumor Recognition from Abdominal CT Images using Transfer Learning

Wasi, Sefatul

Independent University, Bangladesh (IUB)

Draft Version

Kidney Tumor Recognition from Abdominal CT Images using Transfer Learning

Sefatul Wasi
Center for Computational & Data Sciences
Department of Computer Science and Engineering
Independent University, Bangladesh.
Dhaka 1299, Bangladesh.
1821545@iub.edu.bd

Saadia Binte Alam
Center for Computational & Data Sciences
Department of Computer Science and Engineering
Independent University, Bangladesh.
Dhaka 1299, Bangladesh.
saadiabinte@iub.edu.bd

Rashedur Rahman
Graduate School of Engineering
University of Hyogo
2167 Shosha, Himeji, Hyogo, Japan
rashed.riyadh14@gmail.com

M Ashraful Amin
Center for Computational & Data Sciences
Department of Computer Science and Engineering
Independent University, Bangladesh.
Dhaka 1299, Bangladesh.
aminmdashraful@iub.edu.bd

Syoji Kobashi
Graduate School of Engineering
University of Hyogo
2167 Shosha, Himeji, Hyogo, Japan
kobashi@eng.u-hyogo.ac.jp

***Abstract*— Kidney tumor is a health concern that affects kidney cells and may leads to mortality depending on their type. Benign tumors can be unproblematic whereas malignant tumors pose the threat of kidney cancer. Early detection and diagnosis are possible through kidney tumor recognition based on deep learning techniques. In this paper, a method based on transfer learning using deep convolutional neural network (DCNN) is proposed to recognize kidney tumor from computed tomography (CT) images. The proposed method was evaluated on 5284 images. The final accuracy, precision, recall, specificity and F1 score were 92.54%, 80.45%, 93.02%, 92.38% and 0.8628, respectively.**

Keywords—kidney tumor recognition, computed tomography, deep convolutional neural networks, transfer learning.

I. INTRODUCTION

Kidney is one of the major organs that oversee numerous critical functions that sustain the body's chemical equilibrium. Kidney tumors can cause severe complications when it comes to the seamless functioning of the kidneys. Benign and malignant tumors are the two types of kidney tumors that can affect a healthy human being. Although benign tumors are harmless and cause some symptomatic discomfort, malignant ones are the ones that cause kidney cancer [1]. Kidney cancer, also known as renal cancer, is one of the most common cancers worldwide. The number of new cases and fatalities has increased to 400,000 and 180,000 respectively in recent years [2]. It is a significant health concern and can primarily affect mortality rate and the process of prognosis if not detected early. Renal Cell Carcinoma (RCC) is the most common type of renal cancer that covers almost 90% of global cases. Early diagnosis can contribute to increased survival in the case of localized tumors [3]. Unfortunately, some information is still unknown in most kidney tumor cases, whether malignant or benign —from managing localized tumor to handling metastatic kidney cancer. Advancements in imaging techniques contribute towards a preventive approach to improve the treatment of kidney tumors. Practitioners and radiologists cannot manually detect kidney tumors and relevant risk factors with precision. Imaging methods like Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) are widely used for anatomical visualization. However, several strenuous activities have been seen in clinical observations. The precise location of many kidney tumors might be perplexing because of the fluctuation of positioning based on different patients. The diversity of tumor appearances and the texture similarity of tumors and adjacent tissues create complications [4].

The manual kidney tumor detection and segmentation is strenuous and challenging. This is where computer-aided automatic solutions come to play, with significant improvements in providing better treatment. In previous years, many image processing-based solutions emerged as researchers dived deep to understand and conquer the problems related to kidney tumor detection and segmentation. Nader et al. [5] proposed automatic liver tumor segmentation that extracts tumor lesions from CT image slices based on different morphological image processing techniques and knowledge-based constraints for tumor classification. Seo [6] used an approach where automatic hepatic tumor segmentation was executed using sequential steps like segmenting livers and hepatic tumors by utilizing optimal thresholding and provided significantly promising results for small tumors. Lu et al. [7] proposed an active contouring technique to procure tumor boundaries. Chen and Metaxas [8] proposed the use of Markov Random Field (MRF) estimation technique to segment tumors. These techniques have been outdated since the contribution of deep learning models has enabled researchers to get better outcomes in segmentation techniques. Medical image recognition, detection and segmentation combined with deep learning models have been playing a pivotal role that aims to contribute towards a future where the effective diagnosis will overcome current limitations. The recent contribution of CNN and various deep learning models have been immense. Implementation of segmentation, detection, localization, and classification of medical images can provide better clinical success rates and improvement of diagnosis.

Ukai et al. [9] proposed automated pelvic fracture detection on 3-dimensional CT (3D-CT) using multiple 2D object detection systems like YOLOv3. Mahdi et al. [10] proposed teeth recognition using candidate optimization techniques and R-CNN models. Zhao et al. [11] proposed a 3D U-Net based segmentation of kidney and kidney tumors. Shehata et al. [12] utilized deformable model evolution in

conjunction with a Markov-Gibbs random field picture model to propose a shape-based level-set framework for 3D kidney segmentation. Skalski et al. [13] proposed an automatic kidney region classification technique where the identification was based on RUSBoost and decision trees. Cruz et al. [14] used various techniques, starting with AlexNet for scope reduction, 2D U-Net for segmentation and false positive reduction. Lee et al. [15] proposed the detection and segmentation of small renal masses (SRM) in contrast-enhanced CT images.

Different public datasets and challenges have been recently gaining momentum. The 2021 Kidney and Kidney Tumor Segmentation Challenge (KiTS21) is one such challenge that focuses on the automatic semantic segmentation of kidney tumors [16]. Heo [17] proposed an automatic segmentation of kidney and kidney masses, including tumors and cysts, using U-Net based on the KiTS21 grand challenge. Xu et al. [18] proposed a modified version of nnU-Net for the KiTS21 challenge. The notable work of Yang et al. [19] suggests using transfer learning for automatic tumor segmentation based on KiTS21, which showed effective results. Transfer learning (TL) works on implementing knowledge transfer from one related task to another to increase efficiency. Different deep convolutional neural network (DCNN) based approaches require extensive data for training purposes, hence creating data scarcity [20]. This is where TL combined with DCNN provides better results and enhances model effectiveness to overcome new challenges. The idea of parametric knowledge transfer approaches utilizing both TL and DCNN is adopted in this study.

In this paper, we proposed a method based on DCNN and TL for kidney tumor recognition from CT images. CT images containing kidney and tumors are extracted from KiTS21 dataset and pre-processed. Later, the classification layers of a DCNN is trained using the extracted CT images. We used a Visual Geometry Group-16 (VGG16) [21] DCNN pre-trained with ImageNet [22] dataset.

II. METHODOLOGY

Figure 1 provides a comprehensive overview of the proposed approach. At first, CT images with kidney and kidney tumors are extracted and pre-processed. After that the images are used to train a DCNN. Later, the result is evaluated by calculating precision, recall, specificity and F1 score.

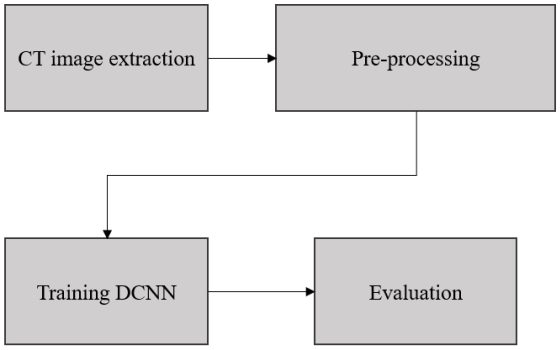

Figure 1: Methodology overview

A. Dataset Description

In this study, KiTS21 dataset is used. Patients in the KiTS21 dataset had partial or radical nephrectomy for suspected kidney cancer between 2010 and 2020 that includes kidney CT images with tumor and without tumor. The KiTS21 dataset is the latest release of the KiTS dataset and holds abdominal CT images of 300 patients. Every patient's data is stored in an anonymized folder, referring to them as cases.

Each case holds the CT data as an 'imaging' file and three independent segmentation files: kidney, tumor, and cyst. We consider the kidney and tumor for our recognition task. So, the two classes are namely- ‘kidney’ and ‘tumor’. ‘Kidney’ class are images of kidney without tumors and ‘tumor’ class are images of kidney with tumors. The imaging file is the abdominal CT scan data in the shape of slice number×height×width where both the height and width of each slice is 512 pixels. We have a corresponding segmentation file for every imaging file with the ground truth labels for the kidney and the tumor. The slice number emulates the axial view of the planes, and the increase in slice number depicts superior to inferior progress of the view. The number of slices ranges from 29 to 1059, with slice thickness ranging from 0.5 to 5 mm and the slice resolution ranging from 0.437 mm per pixel to 1.041 mm per pixel.

At first, the slices of kidney and kidney with tumors are extracted and saved as individual image using the segmentation masks as reference. After that, the images are preprocessed using an active contour algorithm [23] to subtract the background of the CT images. The active contour method is an iteration-based region-growing algorithm for image segmentation. The technique acquires deformable structures of an object present in images utilizing different constraints and factors. Finally, 18,713 images of the kidney class and 7,693 images of the tumor class are extracted. The images are downsized to 224×224 before saving as portable network graphic (png) image.

B. Deep Convolutional Neural Networks

The utilization of VGG16 as DCNN for medical image detection can be crucial and noteworthy. Because of its uniform and simplistic architecture VGG16 is one of the mostly used computer vision architectures for image recognition and classification. This was one of the key reasons for choosing VGG16 for our recognition task. Rather than training a DCNN model from scratch, the use of VGG16 and TL to construct a pre-trained model based on ImageNet offers a more efficient approach to the task.

1) VGG16 Architecture

VGG16 is an iteration of the base VGG with 16 convolution layers and an invariable architecture. The layers have 3×3 small convolution filters. 16 represents the weights of 16 layers in the model. These weights build up the weight layers that are mainly the layers consisting of learnable parameters. Aside from the significant weight layers, the model has thirteen convolutional layers, five maxpooling layers, and three dense layers. The size of the input layer is (224, 224, 3) where 3 represents the 3 color channels of images – red, green, and blue (RGB). The image size is 224×224. The standout feature of VGG16 is the addition of 3x3 convolution layers with a stride of 1. The architecture also utilizes uniform usage of maxpooling layer and padding of 2×2 filters with a stride of 2. The model is structured in a way where we have 5 convolution blocks, each followed by a maxpooling layer. First two blocks have two consecutive convolutional layers where conv-1 has 64 filters and conv-2 has 128 filters. The next three convolution blocks consist of

three consecutive convolutional layers where conv-3 has 256 filters, conv-4 and conv-5 has 512 filters each. The last three fully connected layers (FCL) consisting of 4096 channels in the first two layers. Final FCL contains 1000 channels that contributes to the initial 1000 class classification for ImageNet challenge. The final layer of the model is a softmax activation layer.

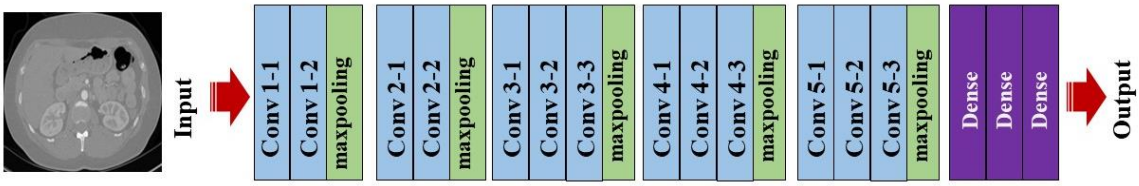

Figure 2: VGG16 network architecture

C. Evaluation Metrics

The parameters used for evaluation are accuracy, precision, recall and F1 score that are calculated from confusion matrix. Confusion matrix is a summarization method of the number of predictions made by a classifier. The basics of a confusion matrix hold the four principal outcomes being true positive (TP), true negative (TN), false positive (FP) and false negative (FN). TP and TN represent successful predictions. FP and FN are considered errors. Using the values obtained by this matrix, we can formalize the equations for calculating accuracy, precision, recall and F1 score [24].

1) Accuracy

Accuracy represents a model's correct prediction performance. The metric evaluates the correctly predicted number of TP and TN for the total predictions and can be described by equation (1).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

2) Precision

Precision evaluates and provides how precisely the model predicts positive labels, TP, out of the total prediction of positives, TP, and FP. So, the equation for calculating precision can be written as equation (2).

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

3) Recall

Recall measures the number of positive observations that are correctly identified as positive. It is the relation between the number of correctly predicted positives, TP, and the total of actual positives, TP, and FN. This relation can be described as equation (3).

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

4) Specificity

Specificity evaluates the relation between TN and the total number of actual negatives. This helps to differentiate between false positive predictions and have an idea about the total number of negatives.

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (4)$$

5) F1 Score

F1 score is the harmonic mean of precision and recall. The use of harmonic mean allows the score to weight accuracy and recall equally and suspends imbalance. It gives us the end value between 0 and 1. The calculation can be derived as equation (5).

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (5)$$

III. Results

We implemented the VGG16 model using ImageNet as the pre-trained weight. A total of 26,406 files consisting of kidney and tumor classes were split in 60:20:20 ratio for training, validation, and test set. The model can classify 1000 distinct labels; however, we only require two. For training, we first load the VGG16 model and freeze training for all convolutional layers. Only the fully connected layers were trained where we set our output features to two. The model was trained for 50 epochs with a learning rate of 0.0001 using the adam optimizer. We added a learning rate scheduler that decreases the learning rate by 0.1 per 7 epochs. Successful training of 50 epochs gave us an accuracy of 92.54% for the test set which consists of 5,284 images. We got precision value of 80.45% and recall value of 93.02% from the model. The specificity and F1 score were 92.38% and 0.8628, respectively. Figure 3 depicts the confusion matrix. Figure 4 shows some examples of model performance on test set.

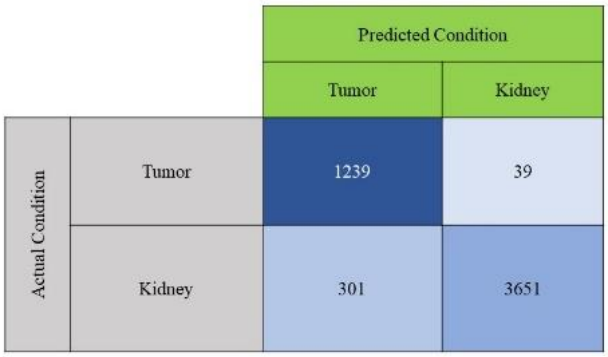

Figure 3: Confusion matrix

IV. Discussion

We calculated the accuracy, precision, recall and F1 score for the test set using the equations discussed in the

methodology. We obtained a precision score of 80.45% in the test set, indicating that 80.45% of the detected tumors were accurately identified out of all recognitions by the trained model.

Recall indicates how many actual true positives or tumors our model was able to find. From the test set of our dataset, 93.02% of the ground truth tumors were correctly recognized by the model.

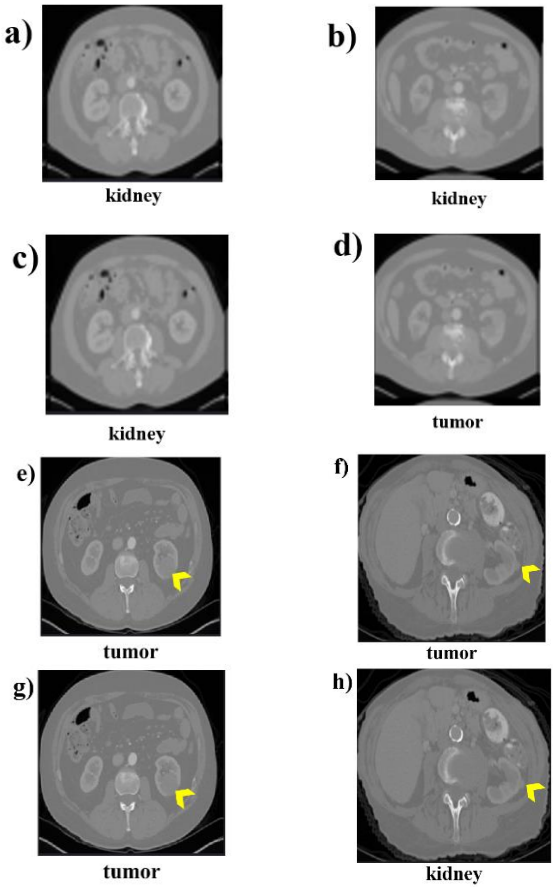

Figure 4: Sample of ground truth and recognition images by VGG16- (a) Ground truth with kidney, (b) Ground truth with kidney, (c) Prediction of a) as kidney, (d) Prediction of b) as tumor, (e) Ground truth with tumor, (f) Ground truth with tumor, (g) Prediction of e) as tumor, (h) Prediction of f) as tumor. The tumor regions are indicated using the yellow marker.

Nevertheless, having a high recall does not necessarily mean we have a good model. Specificity measures the amount of TN which is kidney for our case that are correctly identified. 92.38% of specificity score tells us the number of correctly recognized images with kidney by the model.

The F1 score essentially incorporates the quality and completeness of the recognition in a score that estimates the performance of a model. Relatively good precision of 80.45% interprets that the model recognized the tumors accurately and a higher recall means the model recognized 93.02% of the tumors out of all tumors successfully. This gives us the F1 score of 0.8628 which indicates our model achieved good performance of recognizing the tumors.

Our model has a few shortcomings as well. We only utilized data with kidneys or tumors as images; when the model is supplied with data that does not reflect either of the classes, the performance of the model may change. Another condition that may affect the performance of the proposed method is the presence of implants in the CT images. The presence of implants or any other object can distort the pixel values of a CT image, incapacitating us from processing the images for recognizing kidneys or tumors. The dataset has cysts other than the two classes we are using. Therefore, images with cysts can also prevent us from retaining good results. Another notable drawback is that our model depends on only the images' axial view.

V. Conclusion & Future Works

In this paper, we conducted kidney tumor recognition utilizing the KiTS21 dataset which consists of abdominal CT images of anonymized 300 patients. The goal was to formulate an automatic method to successfully recognize images with tumors. The proposed method used TL approach to leverage the pre-trained VGG16 model and using the CT images as inputs. The parameters may have yielded some commendable results, but there is still room to improve.

The future work includes improving the amount and quality of data by including more images of kidneys, tumors, implants, and other anomalies that can influence the performance of the model to recognize kidney tumors and test the model's capabilities. Another improvement will be the addition of sagittal and coronal views of images.

References

[1] F. Türk, M. Lüy, and N. Barışçı, “Kidney and renal tumor segmentation using a hybrid V-net-based model,” Mathematics, vol. 8, no. 10, p. 1772, 2020.

[2] Globocan. The Global Cancer Observatory. Kidney Fact Sheet; 2020.

[3] Lewis, G., Maxwell, A.: Early diagnosis improves survival in kidney cancer. The Practitioner. 256, 1748, 13-6, 2012.

[4] Q. Yu, Y. Shi, J. Sun, Y. Gao, J. Zhu, and Y. Dai, “Crossbar-net: A novel convolutional neural network for kidney tumor segmentation in CT images,” IEEE Transactions on Image Processing, vol. 28, no. 8, pp. 4060–4074, 2019.

[5] N. H. Abdel-massieh, M. M. Hadhoud, and K. M. Amin, “Automatic liver tumor segmentation from CT scans with knowledge-based constraints,” 2010 5th Cairo International Biomedical Engineering Conference, 2010.

[6] K.-S. Seo, "Automatic Hepatic Tumor Segmentation Using Composite Hypotheses," in Proceedings of the Second International Conference on Image Analysis and Recognition (ICIAR '05), vol. 3656. Toronto, Canada, 2005, pp. 922-929.

[7] R. Lu, P. Marziliano, and C. Hua Thng, "Liver Tumor Volume Estimation by Semi-Automatic Segmentation Method," presented at 27th Annual International Conference of the Engineering in Medicine and Biology Society (EMBS '05), Shanghai, China, 2005.

[8] T. Chen and D. Metaxas, "A hybrid framework for 3D medical image segmentation," Journal of Medical Image Analysis, vol. 9, pp. 547-565, 2005.

[9] K. Ukai, R. Rahman, N. Yagi, K. Hayashi, A. Maruo, H. Muratsu, and S. Kobashi, “Detecting pelvic fracture on 3D-CT using deep convolutional neural networks with multi orientated slab images,” Scientific Reports, vol. 11, no. 1, 2021.

[10] F. P. Mahdi, K. Motoki, and S. Kobashi, “Optimization technique combined with deep learning method for teeth recognition in dental panoramic radiographs,” Scientific Reports, vol. 10, no. 1, 2020.
[11] W. Zhao, D. Jiang, J. Peña Queralta, and T. Westerlund, “MSS U-Net: 3D segmentation of kidneys and tumors from CT images with a multi-scale supervised U-Net,” Informatics in Medicine Unlocked, vol. 19, p. 100357, 2020.
[12] M. Shehata et al., “A level set-based framework for 3D kidney segmentation from diffusion MR images,” 2015 IEEE International Conference on Image Processing (ICIP), 2015.
[13] A. Skalski, J. Jakubowski, and T. Drewniak, “Kidney tumor segmentation and detection on computed tomography data,” 2016 IEEE International Conference on Imaging Systems and Techniques (IST), 2016.
[14] L. B. da Cruz et al., “Kidney segmentation from computed tomography images using Deep Neural Network,” Computers in Biology and Medicine, vol. 123, p. 103906, 2020.
[15] H. S. Lee, H. Hong, and J. Kim, “Detection and segmentation of small renal masses in contrast-enhanced CT images using texture and context feature classification,” *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, 2017.
[16] N. Heller, “The 2021 Kidney Tumor Segmentation Challenge” KiTS21, 2021.
[17] J. Heo, “Automatic segmentation in abdominal CT imaging for the kits21 challenge,” Lecture Notes in Computer Science, pp. 98–102, 2022.
[18] L. Xu, J. Shi, and Z. Dong, “Modified nnU-net for the MICCAI Kits21 Challenge,” Lecture Notes in Computer Science, pp. 22–27, 2022.
[19] X. Yang, J. Zhang, J. Zhang, and Y. Xia, “Transfer learning for KiTS21 challenge,” Lecture Notes in Computer Science, pp. 158–163, 2022.
[20] H. E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros, and T. Ganslandt, “Transfer learning for medical image classification: A literature review,” BMC Medical Imaging, vol. 22, no. 1, 2022.
[21] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv, 2014, [preprint].
[22] Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. "Imagenet: A large-scale hierarchical image database." In 2009 IEEE conference on computer vision and pattern recognition, pp. 248-255. IEEE, 2009.
[23] R. T. Whitaker, A level-set approach to 3D reconstruction from range data. International Journal of Computer Vision, Volume 29, Issue 3, pp. 203-231, 1998.
[24] D. Yang, C. Martinez, L. Visuña, H. Khandhar, C. Bhatt, and J. Carretero, “Detection and analysis of COVID-19 in medical images using Deep Learning Techniques,” Scientific Reports, vol. 11, no. 1, 2021.