Boosting kidney stone identification using two-step transfer learning in endoscopic images

Juan Pablo Betancur-Rengifo[1], Arturo Ruiz-Sanchez[1], Ivan Reyes-Amezcua[1,2], Francisco Lopez-Tiro[1,3,4], Jonathan El-Beze[5], Jacques Hubert[5], Michel Daudon[6], Gilberto Ochoa-Ruiz*[,1,3], and Christian Daul*[,4]

[1] CV-INSIDE Lab Member, Mexico
[2] Centro de Investigación y de Estudios Avanzados, Mexico
[3] Tecnologico de Monterrey, School of Engineering and Sciences, Mexico
[4] CRAN (UMR 7039), Université de Lorraine and CNRS, France
[5] CHU Nancy, Service d'urologie de Brabois, France
[6] Assistance Publique Hôpitaux de Paris, France
*Corresponding authors:
gilberto.ochoa@tec.mx, christian.daul@univ-lorraine.fr

Abstract. Determining the cause of kidney stone formation is crucial to provide adequate treatment and preventing relapses. For this purpose, there are different approaches to visually determining the kidney stone type. However, ex-vivo procedures can take up to weeks, while in-vivo procedures require highly trained specialists. Machine learning models have been developed to assist in the classification of kidney stones in the clinical area, however, they lack data available for training. In this work, a two-step transfer learning approach is used to classify kidney stones. The proposed approach transfers knowledge learned on a set of microscopic images (ex-vivo dataset) to a final model that classifies images from endoscopic images (in-vivo dataset). The results show that learning features from a similar, controlled domain helps to improve the performance of a model that performs classification in real conditions.

Keywords: Kidney stones, Classification, Transfer Learning, Endoscopy

1 Introduction

The formation of kidney stones in the urinary tract is a public health problem [12][14]. In industrialized countries, ten percent of the population suffers an episode of kidney stones. Recent studies have determined that the risk of recurrence increases to 40% in less than 5 years [13][23]. Determining the root cause of kidney stone formation is crucial to avoid relapses through personalized treatment [11][13][26]. Therefore, different approaches for visually identifying some of the most common types (or classes) of kidney stones have been proposed in recent years for application in clinical practice [4][8].

The Morpho-Constitutional Analysis (MCA) is the most important method for identification of the different types on extracted kidney stones (ex-vivo)

[2][10]. MCA consists of two complementary analyses: first, (i) a visual inspection of characteristics (high-quality microscopic images rich in texture, shape, and color) of the surface and section views using a microscope (suitable acquisition conditions and with the possibility of acquiring multiple views of the same stone). Then, (ii) an infrared-spectrophotometry analysis enables to identify the molecular and crystalline composition of the different areas (layers) of the kidney stone [3]. However, the waiting time for the results of an MCA study can take weeks, making it difficult to establish immediate treatment for the patient. On the other hand, removing a complete kidney stone is a difficult procedure. To complicate the issue, in modern endoscopic techniques, kidney stones are pulverized with a laser during surgery. Consequently, the extracted fragments lack of visual information for analysis, and the biochemical composition changes [15].

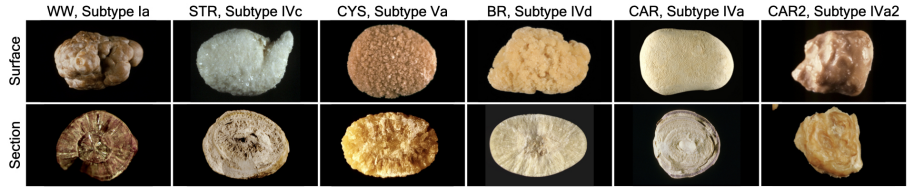
Endoscopic Stone Recognition (ESR) is a technique to determine the type of kidney stone during surgery without the need for biochemical analysis. ESR consists only of visual inspection (is strongly correlated with MCA [10]) on in-vivo endoscopic images through a monitor during the surgical procedure (difficult acquisition conditions and with the constraint of obtaining multiple views of the same kidney stone). However, only a few highly trained experts in this task can determine the type of kidney stone using only endoscopic images. Highly trained experts employ procedures that are time-consuming, subjective, and require intensive training to correctly classify kidney stones [5][16].

In recent years, different methods have been proposed to perform ESR automatically [1][9][17]. These works have shown promising results using techniques based on Deep Learning (DL) methods. However, one of the most common problems in DL-based models for classifying kidney stones in endoscopic images is the lack of data available for the training step. In addition, the similarity of the data distribution is another important factor to obtain an adequate model. Consequently, this suggests a trade-off between the amount of available data and the data distribution to fit the network weights adequately. DL-based models report fine-tuning of weights learned from distributions other than kidney stones (commonly from ImageNet [6]).

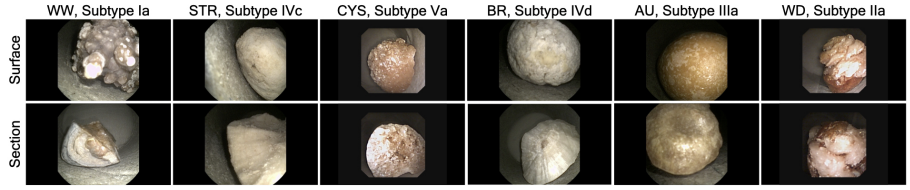
Recently, Transfer Learning has been used in the medical field to use the features learned on one domain (a pre-trained network with millions of images) and leveraging the knowledge to a specific domain (usually performed on tasks where the dataset has too little data to train a complete model from scratch) [21][28]. In the context of kidney stones it is difficult to acquire a large dataset (to train from scratch) from endoscopic images (in-vivo), since the acquisition conditions are adverse (changes in illumination, camera position, particle detachment during surgery, among others). However, due to the nature of the acquisition in microscopic images (ex-vivo) it is easy to capture a similar size dataset but rich in texture, color and shape that can be used to transfer knowledge to an in-vivo imaging model.

In this work, a two-step Transfer Learning model to classify six different types of in-vivo kidney stones is proposed. The model uses homogeneous as well as heterogeneous Transfer Learning phases with a ResNet50 architecture pre-

trained with the ImageNet dataset. In order to validate our proposal, the approach transfers knowledge learned on a small set of microscopic images (ex-vivo dataset rich in color, texture, and color) to a final model that classifies endoscopic images (in-vivo dataset in real-life acquisition conditions). The rest of this paper is structured as follows. Section II gives an overview of the acquired microscopic and endoscopic image dataset and presents the two-step Transfer Learning approach. Section III compares the performances of the two-step Transfer Learning approach with the related work techniques. Finally, Section IV concludes the contribution and proposes perspectives.



(a) Ex-vivo images (microscopic dataset)



(b) In-vivo images (endoscopic dataset)

Fig. 1: Examples of kidney stone images from the (a) ex-vivo images of microscopic dataset [2] and (b) in-vivo images of endoscopic dataset [7]. Each dataset consists of six of the most common classes in kidney stone classification.

2 Materials and Methods

2.1 Datasets

Two different kidney stone datasets were used in our experiments (Table 1). Microscopic images (ex-vivo dataset)[2] and endoscopic images (in-vivo datasets)[7] were acquired and annotated by an expert. The results of visual classification were statistically confirmed by the concordance study in [10]. The dataset's main characteristics are described as follows.

Dataset A. The microscopic images dataset (ex-vivo dataset) [2] consists of 366 microscopic images (Fig. 1a). Two sub-datasets of A for surface and section views consist of 209 and 157 images, respectively. Each sub-dataset A contains six different types sorted by subtype denoted as WW (Whewellite, subtype Ia), CAR (Carbapatite, subtype IVa), CAR2 (Carbapatite, subtype IVa2), STR (Struvite, subtype IVc), BRU (Brushite, subtype IVd) and CYS (Cystine, subtype Va). The images were acquired with digital cameras under controlled lighting conditions, without motion effects, in high quality as well as size, and with uniform background.

Dataset B. The endoscopic images dataset (in-vivo dataset) [7] consists of 409 endoscopic images (Fig. 1b). Two sub-datasets of B for surface and section views consist of 246 and 163 images, respectively. Each sub-dataset B contains six different types sorted by subtype denoted as WW (Whewellite, subtype Ia), WD (Weddellite, subtype, IIa), AU (Acide Urique, subtype IIIa), STR (Struvite, IVc), BRU (Brushite, subtype IVd), and CYS (Cystine, subtype Va). The images of dataset B were captured with an endoscope in a controlled environment. The acquisition conditions are similar to those observed in endoscopy using real extracted stones. For more details see [7].

Table 1: Number of complete microscopic and endoscopic images acquired per dataset.

Urologist	Subtype	Main component	Key	Surface	Section	Total
Michel Daudon (Dataset A)	Ia	Whewellite	WW	50	74	124
	IVa1	Carbapatite	CAR	18	18	36
	IVa2	Carbapatite	CAR2	36	18	54
	IVc	Struvite	STR	25	19	44
	IVd	Brushite	BRU	43	17	60
	Va	Cystine	CYS	37	11	48
	Number of images dataset A			*209*	*157*	*366*
Jonathan El-Beze (Dataset B)	Ia	Whewellite	WW	62	25	87
	IIa	Weddellite	WD	13	12	25
	IIIa	Acide Urique	AU	58	50	108
	IVc	Struvite	STR	43	24	67
	IVd	Brushite	BRU	23	4	27
	Va	Cystine	CYS	47	48	95
	Number of images dataset B			*246*	*163*	*409*

Due to the small size of datasets A and B, as well as class imbalance in each subset, patches were made to increase and balance the number of samples for training and testing. As in previous work, classification is not performed on full images, but using square patches corresponding to the kidney stone in surface or section views (tissues are not observed in the patches). Dataset A and B were manually segmented offline under expert supervision.

As demonstrated in [17][18][20] works, the use of square patches allows sufficient color and texture information to be captured for classification. In addition, the use of patches instead of full surface and section images allows for augmentation and balancing of the datasets. According to [18] a patch size of 256×256 pixels was selected for the A and B datasets. To avoid redundant information between patches of the same kidney stone, there should be an overlap of at least 20 pixels. For each dataset (A and B), a total of 12,000 patches were created for each dataset (1,000 patches per kidney stone type and view). The patches of each dataset were "whitened" using the mean m_i and standard deviation σ_i of the color values I_i^w for each RGB channel with $I_i^w = (I_i – m_i)/\sigma_i$, with i = R, G, B. To avoid data leakage in the datasets, a random, non-repeating dataset partitioning strategy was used. a random, non-repeating dataset partitioning

strategy was used. For the training set 80% was used, while for the test set it was 20%.

2.2 Transfer Learning

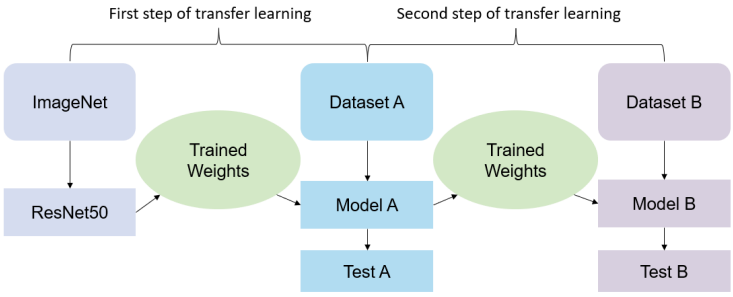


Fig. 2: Two-step Transfer Learning workflow. Model A was initialized with the pre-trained weights of a ResNet50 network and then fine-tuning was applied for Dataset A. Next, Model B starts with the weights learned from Model A which were fine-tuned with Dataset B.

Transfer Learning (TL) is a DL technique whose aim is to use the acquired knowledge of a model trained for a specific domain and apply that model to another domain, relying on the process of domain adaptation [25][**?**]. Besides, depending on the difference between domains, TL can be classified into homogeneous and heterogeneous TL. Homogeneous TL is applicable when there is an available dataset from a domain related to, but not exactly the same, as the target domain. On the other hand, heterogeneous TL is the case where the source and target domains differ [27][**?**]. Furthermore, when a reduced amount of training data is available, it is better to initialize the weights of architecture with pre-trained values rather than random values, i.e., TL from the beginning [22][19]. Due to the difference characteristics and overall size of the datasets in this work, where one was gathered in a more controlled environment than the other one, we propose that the use of a two-step TL heterogeneous approach followed by a homogeneous TL can help in improving the capabilities of a given DL model for classifying kidney stones images from a endoscopic data distribution.

First step of Transfer Learning. Fig. 2 reflects the workflow of the two-step Transfer Learning process. The first step is the heterogeneous phase where the weights are initialized. For this experiment, a ResNet50 architecture pre-trained with ImageNet was used [6] to develop a model capable of classifying the six types of kidney stones from dataset A (subtypes Ia, IVa, IVa2, IVc, IVd, Va). Even if the distribution of the ImageNet dataset differs from the first dataset, as stated above, it is better to initialize the weights this way due to the likeness of the low-level-complexity features from the convolutions [29]. In this step, Gaussian blur and geometrical transformations were applied only to the train images with the aim of giving the model a previous preparation for dataset B. Additionally, a batch size of 24 was used along with a SGD optimizer with

a learning rate of 0.001 and momentum of 0.9, and 30 epochs were executed. Fully connected layers with 768,256,128 and 6 neurons were added with batch-normalization, ReLU activation function, and a dropout of 0.5.

Second step of Transfer Learning. The homogeneous approach was made in the second step of the TL, which consisted in transferring the knowledge (weights) of the trained model from the first step into Dataset B to differentiate between the six types of kidney stones that are in this dataset (subtypes Ia, IIa, IIIa, IVc, IVd, Va). Both steps were addressed to the fine-tuning approach, where all the architecture initializes with pre-trained weights and then those weights are re-trained on the new domain of the required task. The purpose of this methodology is to improve generalization performance and facilitate the finding of good features [24][29]. Here, only geometric transformations were applied because the images in dataset B have acquisition effects (such as blur, illumination changes, among others) that are not observed in dataset A (a controlled acquisition environment). Moreover, 30 epochs were executed also with a SGD optimizer, but with a bigger learning rate of 0.01 since it was expected that the model had less to learn. However, fully connected layers were not added since the idea was to use the trained model without further modifications to the architecture.

Table 2: Mean and standard deviation per metric obtained from each view and step of the transfer learning after 5 executions.

View	Step	Accuracy	Precision	Recall	F1-Score
Surface	1	0.649±0.050	0.655±0.039	0.649±0.050	0.642±0.046
	2	0.832±0.012	0.845±0.012	0.832±0.012	0.829±0.012
Section	1	0.824±0.022	0.834±0.020	0.824±0.022	0.820±0.023
	2	0.904±0.048	0.915±0.037	0.904±0.048	0.903±0.050

3 Results and discussion

Two different experiments were carried out for evaluating the two-step Transfer Learning approach (described in Section 2.2) using patch data (described in Section 2.1). The ability of the two-step TL approach (Fig. 2) to predict different types of kidney stones on endoscopic and microscopic images was evaluated with separate surface and section patches. In Table 2 the well-known metrics (Accuracy, Precision, Recall, and F1-Score) are determined to measure the performance of the model for each view and step.

In the surface view, it can be seen that the weights transferred heterogeneously (different distributions) from ResNet50 (per-trained with ImageNet) to model A (trained on microscopic images) in the first step are a useful strategy to avoid training the model from scratch. Although the performance in test A is relatively low (0.649±0.050 measured with accuracy), in the second step of the surface view, it can be observed that transferring the model A weights to dataset B (trained on endoscopic images) in a homogeneous way (both distributions are similar) improves the performance obtaining an overall accuracy for the B test of 0.832±0.012 (18% compared to test A). The overall increase from step 1 to

step 2 could be due to the high similarity of color characteristics and shapes on the surface of the kidney stone.

In section view, the results obtained for the first step in test A are promising, 0.824±0.022 measured with accuracy. This effect could be due to the fact that the images of dataset A (microscopic images) have rich texture information (concentric structures and layer appreciation) that is impossible to observe in the surface view. In the second step, the accuracy obtained for test B is 0.904±0.048 (an 8% increase over test A). Although the increase from step 1 to step 2 in the section view was not as large with respect to the surface view, surface test B presents the best results obtained in this work.

In order to evaluate the proposed method with respect to The State of the Art (SOTA), Table 3 provides a detailed comparison of the DL-based methods reported for classifying different types of kidney stones using endoscopic image patches. Methods of the related work [1][10][18] were implemented, evaluated on dataset B (endoscopic dataset), and compared to the transfer learning approach described in this work.

Table 3: Comparison between SOTA kidney stones classification methods and the proposed method on test dataset B.

Method	View	Accuracy	Precision	Recall	F1-Score
Martinez, et al. [18]	Surface	0.56±0.23	0.56±0.22	0.56±0.25	0.55±0.23
Black, et al. [1]		0.73±0.19	0.76±0.22	0.73±0.20	0.73±0.18
Estrade, et al. [9]		0.73±0.17	0.75±0.21	0.73±0.19	0.74±0.19
Our Proposal		*0.83±0.01*	*0.84±0.01*	*0.83±0.01*	*0.82±0.01*
Martinez, et al. [18]	Section	0.46±0.12	0.51±0.19	0.46±0.14	0.47±0.13
Black, et al. [1]		0.76±0.18	0.78±0.17	0.76±0.20	0.75±0.16
Estrade, et al. [9]		0.78±0.10	0.80±0.12	0.78±0.11	0.78±0.09
Our Proposal		*0.90±0.04*	*0.91±0.03*	*0.90±0.04*	*0.90±0.05*

The comparative results demonstrate that the two-step transfer learning model for classification six types of kidney stones in endoscopic images proposed in this paper outperforms the implementations found in the SOTA [1][10][18] considering the metrics employed in this research. With this, we can demonstrate that a neural network pre-trained with a general database, tuned with a specific database, and then tuned with a target database, the performance of our network improves without the need to use a large amount of data.

4 Conclusions

In this work, it was demonstrated that it is possible to classify six different types of kidney stones in endoscope images (a small dataset of in-vivo images with real/surgery acquisition conditions) using the knowledge previously learned in a small dataset of microscopic images (ex-vivo images extracted from the patient and with favorable and controlled acquisition conditions). This study confirms that it is easier for a neural network to adjust the weights learned on similar distributions (microscopic images) and adapt them to classify multiple classes in a model of endoscopic images. It is desirable that models of this type should

be improved to classify endoscopic images with the complete kidney stone (as performed by specialists during surgery). We believe that the proposed approach will facilitate training on whole-image models when the datasets are reduced in the number of images.

References

1. Black, K.M., Law, H., Aldoukhi, A., Deng, J., Ghani, K.R.: Deep learning computer vision algorithm for detecting kidney stone composition. BJU international **125**(6), 920–924 (2020)
2. Corrales, M., Doizi, S., Barghouthy, Y., Traxer, O., Daudon, M.: Classification of stones according to michel daudon: a narrative review. European Urology Focus **7**(1), 13–21 (2021)
3. Daudon, M., Dessombz, A., Frochot, V., Letavernier, E., Haymann, J.P., Jungers, P., Bazin, D.: Comprehensive morpho-constitutional analysis of urinary stones improves etiological diagnosis and therapeutic strategy of nephrolithiasis. Comptes Rendus Chimie **19**(11-12), 1470–1491 (2016)
4. Daudon, M., Jungers, P.: Clinical value of crystalluria and quantitative morphoconstitutional analysis of urinary calculi. Nephron Physiology **98**(2), p31–p36 (2004)
5. De Coninck, V., Keller, E.X., Traxer, O.: Metabolic evaluation: who, when and how often. Current Opinion in Urology **29**(1), 52–64 (2019)
6. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
7. El Beze, J., Mazeaud, C., Daul, C., Ochoa-Ruiz, G., Daudon, M., Eschwège, P., Hubert, J.: Evaluation and understanding of automated urinary stone recognition methods. BJU international (2022)
8. Estrade, V., Daudon, M., Traxer, O., Meria, P.: Why should urologist recognize urinary stone and how? the basis of endoscopic recognition. PROGRES EN UROLOGIE **27**(2), F26–F35 (2017)
9. Estrade, V., Daudon, M., Richard, E., Bernhard, J.c., Bladou, F., Robert, G., Denis de Senneville, B.: Towards automatic recognition of pure and mixed stones using intra-operative endoscopic digital images. BJU international **129**(2), 234–242 (2022)
10. Estrade, V., Denis de Senneville, B., Meria, P., Almeras, C., Bladou, F., Bernhard, J.C., Robert, G., Traxer, O., Daudon, M.: Toward improved endoscopic examination of urinary stones: a concordance study between endoscopic digital pictures vs microscopy. BJU international **128**(3), 319–330 (2021)
11. Friedlander, J.I., Antonelli, J.A., Pearle, M.S.: Diet: from food to stone. World journal of urology **33**(2), 179–185 (2015)
12. Hall, P.M.: Nephrolithiasis: treatment, causes, and prevention. Cleveland Clinic journal of medicine **76**(10), 583–591 (2009)
13. Kartha, G., Calle, J.C., Marchini, G.S., Monga, M.: Impact of stone disease: chronic kidney disease and quality of life. Urologic Clinics **40**(1), 135–147 (2013)
14. Kasidas, G., Samuell, C., Weir, T.: Renal stone analysis: why and how? Annals of clinical biochemistry **41**(2), 91–97 (2004)
15. Keller, E.X., De Coninck, V., Audouin, M., Doizi, S., Bazin, D., Daudon, M., Traxer, O.: Fragments and dust after holmium laser lithotripsy with or without "moses technology": How are they different? (2019)
16. Krambeck, A.E., Lingeman, J.E., McAteer, J.A., Williams, J.C.: Analysis of mixed stones is prone to error: a study with us laboratories using micro ct for verification of sample content. Urological research **38**(6), 469–475 (2010)
17. Lopez, F., Varelo, A., Hinojosa, O., Mendez, M., Trinh, D.H., ElBeze, Y., Hubert, J., Estrade, V., Gonzalez, M., Ochoa, G., et al.: Assessing deep learning methods for the identification of kidney stones in endoscopic images. In: 2021 43rd Annual

International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). pp. 2778–2781. IEEE (2021)
18. Martínez, A., Trinh, D.H., El Beze, J., Hubert, J., Eschwege, P., Estrade, V., Aguilar, L., Daul, C., Ochoa, G.: Towards an automated classification method for ureteroscopic kidney stone images using ensemble learning. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). pp. 1936–1939. IEEE (2020)
19. Ng, H.W., Nguyen, V.D., Vonikakis, V., Winkler, S.: Deep learning for emotion recognition on small datasets using transfer learning. ICMI 2015 - Proceedings of the 2015 ACM International Conference on Multimodal Interaction pp. 443–449 (2015). https://doi.org/10.1145/2818346.2830593
20. Ochoa-Ruiz, G., Estrade, V., Lopez, F., Flores-Araiza, D., Beze, J.E., Trinh, D.H., Gonzalez-Mendoza, M., Eschwège, P., Hubert, J., Daul, C.: On the in vivo recognition of kidney stones using machine learning. arXiv preprint arXiv:2201.08865 (2022)
21. Raghu, M., Zhang, C., Kleinberg, J., Bengio, S.: Transfusion: Understanding transfer learning for medical imaging. Advances in neural information processing systems **32** (2019)
22. Raghu, M., Zhang, C., Kleinberg, J., Bengio, S.: Transfusion: Understanding transfer learning for medical imaging. In: Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 32. Curran Associates, Inc. (2019), https://proceedings.neurips.cc/paper/2019/file/eb1e78328c46506b46a4ac4a1e378b91-Paper.pdf
23. Scales Jr, C.D., Smith, A.C., Hanley, J.M., Saigal, C.S., in America Project, U.D., et al.: Prevalence of kidney stones in the united states. European urology **62**(1), 160–165 (2012)
24. Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., Liu, C.: A survey on deep transfer learning. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) **11141 LNCS**, 270–279 (2018). https://doi.org/10.1007/978-3-030-01424-7-27
25. Torrey, L., Shavlik, J.: Transfer learning. In: Handbook of research on machine learning applications and trends: algorithms, methods, and techniques, pp. 242–264. IGI global (2010)
26. Viljoen, A., Chaudhry, R., Bycroft, J.: Renal stones. Annals of clinical biochemistry **56**(1), 15–27 (2019)
27. Weiss, K., Khoshgoftaar, T.M., Wang, D.D.: A survey of transfer learning, vol. 3. Springer International Publishing (2016). https://doi.org/10.1186s40537-016-0043-6
28. Wen, Y., Chen, L., Zhou, C., Deng, Y., Zeng, H., Xi, S., Guo, R.: On the effective transfer learning strategy for medical image analysis in deep learning. In: 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 827–834. IEEE (2020)
29. Yosinski, J., Clune, J., Bengio, Y., Lipson, H.: How transferable are features in deep neural networks ? **27** (2014)