2021 International Conference on Forthcoming Networks and Sustainability in AIoT Era (FoNeS-AIoT) | 978-1-6654-1091-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/FoNeS-AIoT54873.2021.00010

Performance Evaluation for Various Deep Learning (DL) Methods Applied to Kidney Stone Diseases

Özlem Sabuncu
Department of Electrical and Electronic Engineering
Near East University
Nicosia, Mersin 10, Turkey
ozlem.sabuncu@neu.edu.tr

Bülent Bilgehan
Department of Electrical and Electronic Engineering
Near East University
Nicosia, Mersin 10, Turkey
bulent.bilgehan@neu.edu.tr

Abstract— **Kidney stone disease is a urological system problem with an increased lifetime prevalence. The formation of crystalline stones in the kidneys reduces kidney function. Different imaging methods are used to detect kidney stones. These images are analyzed by clinical experts. Various methods are developed to assist specialists in diagnosis. The increased number of cases in the healthcare system led to apply artificial intelligence (AI) as a new technological tool to improve time delay management. There are several DL algorithms applied to determine kidney stones. The prime object is to identify the kidney stone at any size under any circumstances using a minimum number of computed tomography (CT) images at the highest accuracy. It is for this reason that this paper classifies the most applicable DL algorithm to be selected with high confidence. The research uses real recorded CT images for evaluation. The Inception-V3 model has the highest test accuracy of 98.52%.**

Keywords—Kidney stone, computed tomography, artificial intelligence, deep learning, convolutional neural network

I. Introduction

The accumulation of acid salts and minerals in the kidneys is known as kidney stone disease [1]. Kidney stone disease, which is thought to have increased in the last few decades, is a common problem affecting 1 in 10 people. Imaging techniques for kidney stone disease play an important role in diagnosis, treatment, or follow-up after urological interventions. Unenhanced computed tomography (CT) is the preferred method for the evaluation of kidney stones because it provides specificity and sensitivity of over 95% for the diagnosis of the disease [2].

Radiologists need to review multiple CT scans to diagnose and evaluate kidney stone disease, but evaluating each patient is time-consuming. The development of an automated classifier instead of these procedures can provide technical support to radiologists and relieve the burden of radiologists [3].

Deep learning methods have promising potential in detecting and evaluating different diseases, as they perform adequately in assigned tasks compared to human experts. Deep neural networks use tags for feature extraction from complex data used for these tasks [4]. Deep learning-based convolutional neural networks (CNNs) are trained with data labeled in relevant classes and can learn the relationship between these class labels. Sufficient training data is required for the successful training of CNNs. CNN models previously trained with the ImageNet dataset with 1000 class labels can be used to initiate a different task using the learned parameters, the transfer learning method [5].

The application of the new technological tool of AI is extended to aid different diagnoses in the healthcare system. The paper [6] applied a neural network to improve the identification of Parkinson's disease at early stages. The author has identified the best-performing algorithm that can be applied with great success.

The paper [7] applied AI to predict diabetic disease that is considered to be one of the great threats to human life. The authors identified Random Forest (RF) classifier to have the best accuracy amongst the others. The success of RF can be applied to form a healthcare network based on the internet of things (IoT).

The parallel technological improvement in the field of smartphones has increased the demand for AI applications. The use of AI provided an advantage on the time consumption for disease diagnosis and distributed to a large number of people using the services offered by the IoT network [8].

This study performs the classification of kidney stones with the transfer learning method using 5 different pre-trained CNN models and the performance comparison of the models in the classification task.

II. Materials and Methods

The data used in this study were approved by the ethics committee of the hospital where the data were obtained. Abdominal CT scans of 120 patients from the Dr. Burhan Nalbantoğlu Hospital CT database were collected for this study. 681 CT sections obtained from 120 collected patients were labeled in two separate classes by the radiologist as with and without kidney stones. In this way, the dataset was organized as sections 216 with kidney stones and 465 without kidney stones. Labeled data is reserved separately, 20% for testing, 20% for validation, and the rest for training.

To reduce the overfitting effect, images reserved for training and validation have been updated using the cropping

978-1-6654-1091-5/21/$31.00 ©2021 IEEE
DOI 10.1109/FoNeS-AIoT54873.2021.00010

method. At the same time, the dataset was developed using data augmentation methods such as rotation and flip. These procedures were carried out separately for sections with and without kidney stones. The details of our dataset are presented in Fig. 1.

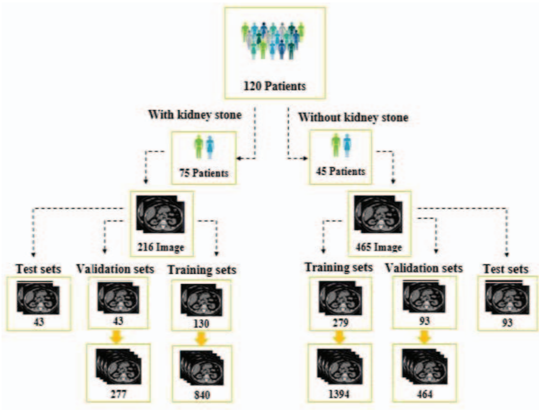

Figure 1. Dataset description

Deep learning-based CNNs are good at extracting valuable features from input data in tasks such as medical disease diagnosis [9].

The focus of the 5 different CNNs used here is to detect kidney stones from the CT images, thus classifying the CT images as with kidney stones or without kidney stones. The CNNs selected for this study, which were shown as the most successful in the performance ranking; Inception-V3 [10], InceptionResNet-V2 [11], Xception [12], NasNet-Mobile [13], and DenseNet-201 [14, 15].

Matlab software was used to train, validate and test the networks used. Before our dataset is trained, the input images are resized for each model separately; 299x299 for Xception, InceptionResNet-V2, and Inception-V3, 224x224 for NasNet-Mobile and DenseNet-201. The training and testing phases are shown in Fig. 2.

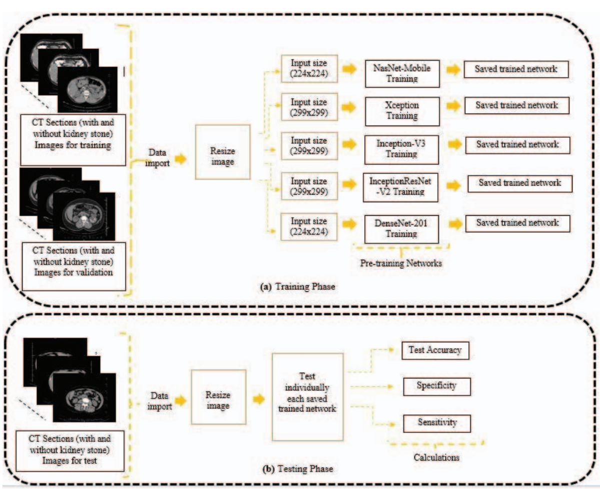

Figure 2. Block diagram of the proposed system, a Training phase, where the training and validation set will be used to train the five networks. b Testing phase where the trained network set will be used for performance evaluation as in Fig. 2

Then, the parameters of each model used were fine-tuned and retrained with our dataset. These processes are called transfer learning. An initial learning rate of 0.0085 and training adjustments of 7 minibatch size were made for all layers of the network. The training results of each model were recorded separately to be used in the testing phase. From the test results; True-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN) values were obtained and recorded. Performance evaluations were made according to the values obtained after these procedures.

To evaluate the performance of the trained models; test accuracy (i.e., Test acc = ((TP+TN)/(TP+FP+TN+FN)), sensitivity (i.e., Sens = TP/(TP+FN)), and specificity (i.e., Spec = FN/(FN+FP)) parameters were considered and calculated. These parameter values and calculation results are presented in Table I.

TABLE I. RESULTS OF TESTS AND CALCULATIONS (UNIT: %)

NWK	TP	FN	FP	TN	SP	SP	TA
Inception-V3	41	2	0	93	95,34	100,00	98,52
Inception ResNet-V2	34	9	11	82	79,06	88,17	85,29
NasNet-Mobile	37	6	4	89	86,04	95,69	92,64
Xception	28	15	29	64	65,11	68,81	67,64
DenseNet -201	29	14	25	68	67,44	73,11	71,32

The abbreviated headings are listed below
NWK: Network
TP: True-positive
FN: False-negative
FP: False-positive
TN: True-negative
SN: Sensitivity
SP: Specificity
TA: Test accuracy

III. RESULT

The deep learning process offers an extensive number of algorithms to apply. It is important to know the best performing algorithm in the application. It şs for thşs reason that classical deep learning algorithms are selected and trained with real CT images. After the training process, the algorithm identifies the stone at the higest accuracy level. In this test, we applied five convolutional neural networks on kidney stones disease for classification purposes. The results of the experiments are shown in Table I.

The performance of the five selected CNNs is evaluated for classification with and without kidney stones. For the performance criteria of the models, TP, TN, FP, FN data should be evaluated together. These evaluation results are presented in Fig. 3.

As can be seen in Fig. 3, the Inception-V3 model achieved the highest test accuracy, sensitivity, and specificity value compared to other models. Then, NasNet-Mobile fulfilled the assigned classification task in the second success. In the performance comparison of five CNN, the Xception model achieved the lowest success.

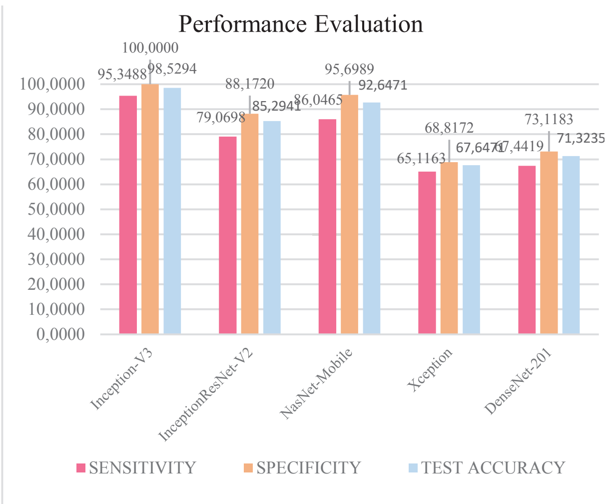

Figure 3. Performance evaluation of the models under study based on the sensitivity, specificity, and test accuracy

IV. Conclusion

In this article, a limited number of CT images were used from a non-public database. The process uses training and testing phases with CNNs to classify under two categories, with and without kidney stones. The work compares the performance of the five different CNN models.

Presented in this study automatic classification of kidney stone disease with deep learning can provide technical support to radiologists as well as assisting the medical specialists. The study also reflects the advantage of such a method where there is no expert access.

It has been observed that the performance evaluation of the five models is in harmony with the radiologist's performance.

The test accuracy, sensitivity, and specificity of the Inception-V3 model achieved the highest performance as 98.52%, 95.34%, and 100.00% respectively.

References

[1] Waldman, S. D. (2019). Atlas of Common Pain Syndromes E-Book. Elvesier Health Sciences.

[2] McCarthy, C. J., Baliyan, V., Kordbacheh, H., Sajjad, Z., Sahani, D., & Kambadakone, A. (2016). Radiology of renal stone disease. *International Journal of Surgery*, *36*, 638-646.

[3] Cui, Y., Sun, Z., Ma, S., Liu, W., Wang, X., Zhang, X., & Wang, X. (2021). Automatic Detection and Scoring of Kidney Stones on Noncontrast CT Images Using STONE Nephrolithometry: Combined Deep Learning and Thresholding Methods. *Molecular Imaging and Biology*, *23*(3), 436-445.

[4] Keles, A., Keles, M. B., & Keles, A. (2021). COV19-CNNet and COV19-ResNet: diagnostic inference Engines for early detection of COVID-19. *Cognitive Computation*, 1-11.

[5] Brinker, T. J., Hekler, A., Utikal, J. S., Grabe, N., Schadendorf, D., Klode, J., ... & Von Kalle, C. (2018). Skin cancer classification using convolutional neural networks: systematic review. *Journal of medical Internet research*, *20*(10), e11936.

[6] Rasheed, J., Hameed, A. A., Ajlouni, N., Jamil, A., Özyavaş, A., & Orman, Z. (2020, October). Application of adaptive back-propagation neural networks for Parkinson's disease prediction. In *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)* (pp. 1-5). IEEE.

[7] Yahyaoui, A., Jamil, A., Rasheed, J., & Yesiltepe, M. (2019, November). A decision support system for diabetes prediction using machine learning and deep learning techniques. In 2019 1st International Informatics and Software Engineering Conference (UBMYK) (pp. 1-4). IEEE.

[8] Rasheed, J., Alimovski, E., & Rasheed, A. (2019, November). mManagement: Wi-Fi Hotspot based Attendance Application using Android Smartphone. In 2019 1st International Informatics and Software Engineering Conference (UBMYK) (pp. 1-5). IEEE.

[9] Rahimzadeh, M., & Attar, A. (2020). A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2. *Informatics in Medicine Unlocked*, *19*, 100360.

[10] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818-2826).

[11] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017, February). Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-first AAAI conference on artificial intelligence*.

[12] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1251-1258).

[13] Zoph, B., Vasudevan, V., Shlens, J., & Le, Q. V. (2018). Learning transferable architectures for scalable image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8697-8710).

[14] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).

[15] Chollet F, others. Keras [Internet]. GitHub; 2015. Available from: https://keras.io/api/applications/