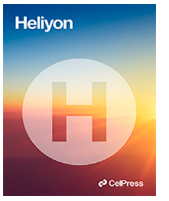
Heliyon 10 (2024) e31255

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

# Computer-assisted decision support for the usage of preventive antibacterial therapy in children with febrile pyelonephritis: A preliminary study

Zhengguo Chen [a], Ning Li [b,*], Zhu Chen [a], Li Zhou [a], Liming Xiao [a], Yangsong Zhang [a,b,c,**]

[a] NHC Key Laboratory of Nuclear Technology Medical Transformation (MIANYANG CENTRAL HOSPITAL), Mianyang, 621000, China
[b] School of Computer Science and Technology, Laboratory for Brain Science and Medical Artificial Intelligence, Southwest University of Science and Technology, Mianyang, 621010, China
[c] Key Laboratory of Testing Technology for Manufacturing Process, Ministry of Education, Southwest University of Science and Technology, Mianyang, 621010, China

ARTICLE INFO

*Keywords:*
Urinary tract infection
Deep learning
Convolutional neural networks
Classification

ABSTRACT

Urinary tract infection (UTI) is one of the most common infectious diseases among children, but there is controversy regarding the use of preventive antibiotics for children first diagnosed with febrile pyelonephritis. To the best of our knowledge, no studies have addressed this issue by the deep learning technology. Therefore, in the current study, we conducted a study using $^{99m}$Tc – DMSA renal static imaging data to investigate the need for preventive antibiotics on children first diagnosed with febrile pyelonephritis under 2 years old. The self-collected dataset comprised 64 children who did not require preventive antibiotic treatments and 112 children who did. Using several classic deep learning models, we verified that it is feasible to screen whether the first diagnosed children with febrile pyelonephritis require preventive antibacterial therapy, achieving a graded diagnosis. With the AlexNet model, we obtained accuracy of 84.05%, sensitivity of 81.71% and specificity of 86.70%, respectively. The experimental results indicate that deep learning technology could provide a new avenue to implement computer-assisted decision support for the diagnosis of the febrile pyelonephritis.

## 1. Introduction

Urinary tract infection (UTI) is a common clinical infectious disease and cause of fever in children [1]. The overall prevalence of urinary tract infection in children with fever younger than 2 years old in the emergency ward is as high as 14% [2]. Around 30% of infants and young children will suffer from recurrent UTI within 6-12 months after their initial infection [3]. Pyelonephritis is one of the common types of UTI. Pyelonephritis is associated with the formation of renal scars, and each recurrence increases the risk of renal scar formation by 2.8% [4]. The primary goal in the management of a child with pyelonephritis is to prevent recurrence of pyelonephritis and acquired renal damage. Although the evidence in favor of antibiotic prophylaxis remains doubtful in

* Corresponding author.
** Corresponding author at: NHC Key Laboratory of Nuclear Technology Medical Transformation (MIANYANG CENTRAL HOSPITAL), Mianyang, 621000, China.
*E-mail addresses:* liningacademy@163.com (N. Li), zhangysacademy@gmail.com (Y. Zhang).

https://doi.org/10.1016/j.heliyon.2024.e31255
Received 13 November 2023; Received in revised form 13 May 2024; Accepted 13 May 2024

Available online 16 May 2024

preventing renal scars associated with VUR, it remains the first line treatment for high-grade reflux and recurrence of pyelonephritis. Approximately 10% – 15% of children will eventually have irreversible kidney scars, which can lead to various complications such as hypertension, proteinuria, and end-stage renal disease [5]. Accurate assessment of febrile pyelonephritis helps to take effective intervention measures as early as possible, reducing the risk of renal scarring and other complications. Therefore, early diagnosis, prevention of recurrent pyelonephritis in children, and appropriate treatment management are crucial for optimal outcomes [6].

Imaging evaluation methods for children with pyelonephritis include renal-bladder ultrasound (RBUS), $^{99m}$Tc – DMSA renal static scanning, and voiding cystourethrogram (VCUG), etc. Although the imaging evaluation method for children with pyelonephritis is still controversial, the top-down or bottom-up imaging evaluation recommendations indicate that scanning is the gold standard for the diagnosis of pyelonephritis or renal scar, which is helpful to confirm or exclude the above diagnosis results [7]. In addition, $^{99m}$Tc – DMSA scanning is more valuable than ultrasound in predicting vesicoureteral reflux (VUR). Therefore, by observing the $^{99m}$Tc – DMSA scanning results, clinicians can decide whether to execute VCUG [5,8]. Furthermore, $^{99m}$Tc – DMSA scanning can investigate the progress of renal injury, from the initial acute pyelonephritis damage to irreversible renal function loss [9].

$^{99m}$Tc – DMSA scanning improves the accuracy of diagnosis and staging of pyelonephritis in children, enhances personalized treatment strategy, and improves the ability to predict outcomes. However, traditional nuclear medicine imaging, including $^{99m}$Tc – DMSA scan, has the characteristics of low spatial resolution and high background noise, which reduces the contrast between lesion tissue and background [10]. In addition, visual evaluation on images can be subjective and may result in differences among different judges. Furthermore, there may be differences in the tracer distribution in different regions of the kidney, making visual evaluation unable to accurately evaluate lesions with low contrast [11]. In the past decade, the emergence of deep learning has provided new avenues for medical image analysis. Deep learning can overcome the main problem of iterative image reconstruction based on traditional models of nuclear medicine and shows broad prospects for image reconstruction, image synthesis, differential diagnosis generation, and treatment guidance [12]. Recent studies, such as Chaudhary et al. [13], have demonstrated the efficacy of deep learning algorithms in denoising $^{99m}$Tc – DMSA scans. In addition, the work of Lee et al. has been instrumental in demonstrating the predictive power of deep learning models for assessing the risk of recurrent pyelonephritis using $^{99m}$Tc – DMSA scans [14]. Thus, deep learning technology can serve as a powerful tool to improve the accuracy and efficiency of clinical decision-making. Leveraging $^{99m}$Tc – DMSA imaging data, computer technologies and artificial intelligence algorithms can assist the clinicians with the personalized diagnosis and treatment of pyelonephritis in children.

During the diagnosis process of pyelonephritis, controversy exists regarding the use of preventive antibiotics in children with febrile pyelonephritis. To address this issue, we proposed a strategy that using the $^{99m}$Tc – DMSA renal static imaging data to investigate whether to use preventive antibiotics on children first diagnosed with febrile pyelonephritis under 2 years old. We collected a dataset comprising 64 children who did not require preventive antibiotic treatments and 112 children who did. We used several classic deep learning models to verify the feasibility of the proposed solution. Interestingly, we obtained an accuracy of 84.05%, a sensitivity of 81.71%, and a specificity of 86.70%, respectively, in the two-class classification problem. This study may offer a potential solution to the controversy surrounding the use of preventive antibiotics in children first diagnosed with febrile pyelonephritis.

## 2. Method and materials

In the following sections, we will introduce the data processing, the DL models and the implementation in detail. Data pre-processing was conducted to reduce the background noise in original images, and data augmentation was used to alleviate the problem of limited sample sizes.

### *2.1. Dataset*

The parents of the children were asked to read and sign an informed consent form before the children participated in SPET image scanning, according to the Declaration of Helsinki (1991). This study has obtained approval from the Biomedical ethics committee of Mianyang Centre Hospital. The approval was granted on 2023, with the reference number S20230216-01. The study's ethical review process ensures compliance with standards and protects participants' rights. In total, 176 children first diagnosed with febrile pyelonephritis were recruited to participate in SPET image scanning, i.e., $^{99m}$Tc – DMSA renal static imaging. The inclusion criteria were as follows: (1) Age $\leq$ 24 months; (2) Children with fever $\geq$ 38.3 °C, elevated C-reactive protein, abnormal urine sediment (pyuria, bacteriuria, or hematuria) or positive urine culture (single bacterial colony number $\geq$ 105 /ml); (3) $^{99m}$Tc – DMSA scan with abnormal results, and CeVUS examination performed within one week after infection control. The exclusion criteria were as follows: (1) Failure to undergo CeVUS or DMSA examination, or no abnormalities in DMSA examination results. (2) Secondary vesicoureteral reflux, such as neurogenic bladder, posterior urethral valve, etc. (3) Children with other urinary tract diseases. Among these children, 64 children were diagnosed to not require preventive antibiotic treatments, and the remaining 112 children were diagnosed to require the treatments.

Intravenous injection of $^{99m}$Tc – DMSA at a dose of 0.05 mCi/kg, was performed 2-3 hours after injection for conventional image acquisition, using a Siemens SPECT machine (model Symbia T6). The scanning parameters were set as follows: low-energy and high resolution collimator, energy peak of 140 keV, window width of 20%, acquisition matrix of 512 × 512, acquisition matrix of 1.78, and acquisition count of 1000K. The children lay in a supine position, with the detector field of view covering the abdominal cavity and pelvic cavity, and posterior, anterior, left posterior oblique, and right posterior oblique positions were acquired, respectively.

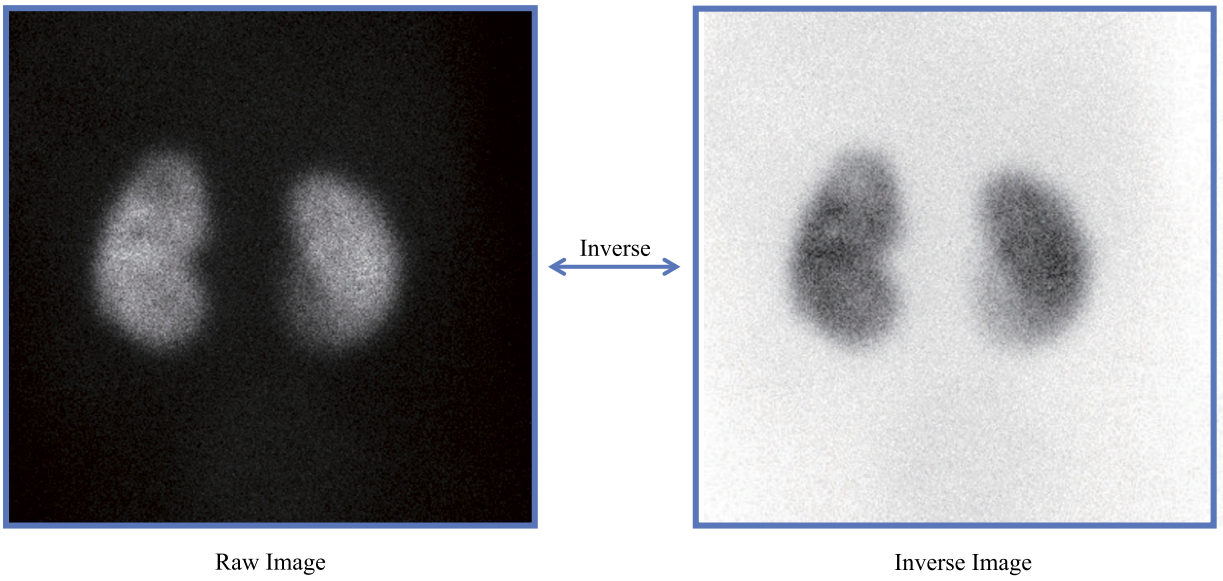



**Fig. 1.** The visualization of noise in the background of $^{99m}$Tc – DMSA images. An inversed image is provided to enhance the visibility of noise within the visual representation.

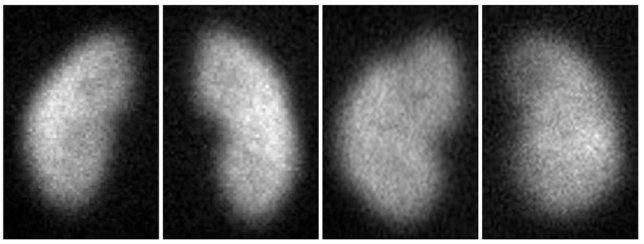


**Fig. 2.** The $^{99m}$Tc – DMSA images of normal left and right kidneys, and the abnormal left and right kidneys are displayed from left to right, respectively.

Compared to adults, children are generally more sensitive to radiation. Therefore, it is crucial to strike a balance between the radiation risk associated with $^{99m}$Tc – DMSA renal static scans and the necessity of this examination. In this study, pediatric patients undergoing $^{99m}$Tc – DMSA renal static scans strictly adhered to the guidelines set by the Pediatric Nephrology Group of the Chinese Medical Association. Simultaneously, the recommended pediatric dosage from nuclear medicine guidelines was meticulously followed, minimizing the administered radioactive drug activity and thereby reducing radiation exposure. In practice, the radiation dose from $^{99m}$Tc – DMSA renal scans in infants and children is exceedingly low ($<0.7$ mSv) [15]. Selçuk Varol et al. assessed 27 pediatric patients and found that the effects of $^{99m}$Tc – DMSA scintigraphy were insufficient to cause oxidative damage. Even though it could induce DNA damage directly through ionizing radiation, such damage is reparable within a short timeframe [16].

### 2.2. Data pre-processing

$^{99m}$Tc – DMSA images often contain a considerable amount of background noise caused by non-uniform isotope distribution and involuntary patient motion [17]. Fig. 1 showcases a substantial amount of noise surrounding the kidneys. The accuracy of neural network inference would be adversely affected by such external noise. To tackle this problem, a 'crop' operation is employed. Rather than using an image with both kidneys for prediction, we cropped the images in half such that the left and right kidneys are predicted independently. Furthermore, due to differences in physical conditions and varying severity of kidney damage among patients, we retained cropped images based on different situations. Specifically, for normal kidneys, the left and right kidneys are preserved separately. For unilateral renal lesions, only the diseased side of the kidney was retained, while the normal side of the kidney was removed. For bilateral renal lesions, both sides were preserved as with normal kidneys. After data pre-processing, the background noise is substantially minimized in the new images. Moreover, the number of normal samples increased to match the number of abnormal samples, alleviating the problem of imbalanced sample sizes between the two categories in the original dataset. Four processed $^{99m}$Tc – DMSA images of normal and abnormal kidneys are shown in Fig. 2. All training and test images were resized to $224 \times 336$ as input to the model.

Training a deep learning model usually requires a large amount of training data to optimize the model parameters [18]. The model is prone to overfitting with insufficient sample data [19]. However, the size of medical imaging data is often limited due to

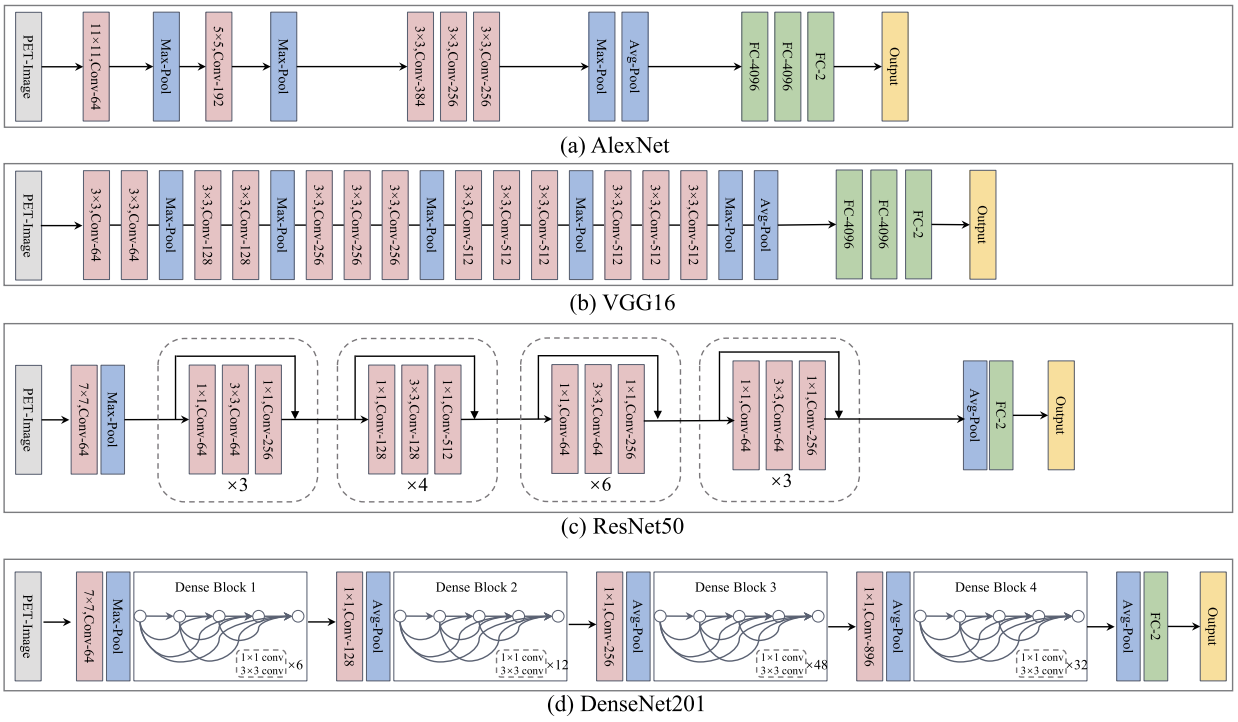



**Fig. 3.** The architecture diagrams of the four models used in this study. (a) AlexNet, (b) VGG16, (c) ResNet50, (d) DenseNet201.

lengthy accumulation periods or concerns over patient privacy. To this end, data augmentation was employed in this study to expand the dataset size to alleviate overfitting problems [20,21]. Considering the mirror symmetry between the left and right kidneys, the training set is expanded by using horizontal flip for the images. The same data augmentation technique was applied to both normal and abnormal kidney images. It should be noted that data augmentation was applied only to the training set, not the testing set, in subsequent experiments.

### 2.3. DL models

Effective feature extraction from medical images is essential to improve classification performance [22]. In traditional computer vision (CV) tasks, extracting features from images relies on CV engineers' experience and long-term experimentation, which is time-consuming and labor-intensive [23]. Compared with manual feature extraction, CNN can automatically extract effective features for classification from images through end-to-end learning. In this study, we aim to investigate the feasibility of predicting whether the children first diagnosed with febrile pyelonephritis need to use preventive antibiotics during the treatment. Therefore, we leveraged four classic CNN models to classify $^{99m}$Tc – DMSA images, i.e., two shallower network models such as AlexNet [24] and VGG16 [25], and deeper network models such as ResNet50 [26] and DenseNet201 [27]. The structures of these networks are depicted in Fig. 3 (a)-(d).

At the 2012 ImageNet large scale visual recognition challenge (ILSVRC), Alexnet achieved breakthrough results compared to the conventional computer vision techniques. It is a typical feedforward CNN model, mainly consisting of five convolutional layers and three fully connected layers [24]. Other components, such as Rectified Linear Unit (ReLU) activation functions [28], max pooling, and dropout [19], are nested within these layers, providing the potential to suppress overfitting and improve the network's learning ability. AlexNet is shown in Fig. 3(a).

Compared with AlexNet, VGG16 has a deeper network structure, which enables it to achieve better performance on the ImageNet [25]. As shown in Fig. 3(b), VGG16 uses smaller convolution kernel of size 3 × 3 instead of large convolution kernel in AlexNet to achieve higher computational efficiency.

Different from AlexNet and VGG16, ResNet [26] and Densenet [27] accomplish very deep network architectures via residual connections and dense connections, respectively. As shown in Fig. 3(c), ResNet utilizes shortcut connections between different layers in the network to learn residual functions, easing the training of deeper network models. As shown in Fig. 3(d), differing from ResNet, DenseNet allows each layer to receive information from all previous layers, thus achieving dense connectivity. Dense connections allow the feature-maps of all preceding layers directly serve as the input into all subsequent layers. In this paper, we used ResNet50 with 50 layers and DenseNet201 with 201 layers to conduct the $^{99m}$Tc – DMSA image classification experiment.

**Table 1**
The classification results of four models. Six evaluation metrics were calculated. The best result under each metric is highlighted in bold black. * and ** indicate the statistically significant differences of $p \leq 0.05$ and $p \leq 0.01$ by the paired *t*-test, respectively.

| Method | Metric | | | | | |
|---|---|---|---|---|---|---|
| | Accuracy (%) | AUC | PPV (%) | NPV (%) | Sensitivity (%) | Specificity (%) |
| AlexNet | **84.05 ± 0.95** | **84.21 ± 0.93** | **87.79 ± 1.81** | **81.90 ± 1.45** | **81.71 ± 2.10** | **86.70 ± 2.27** |
| VGG16 | 82.15 ± 0.87** | 82.25 ± 0.93** | 85.30 ± 1.55* | 80.22 ± 1.18* | 80.21 ± 1.57 | 84.28 ± 2.05 |
| ResNet50 | 75.43 ± 2.52** | 75.50 ± 2.47** | 80.68 ± 3.76** | 73.32 ± 2.45** | 72.64 ± 2.86** | 78.37 ± 5.17** |
| DenseNet201 | 81.20 ± 2.53** | 81.17 ± 2.63** | 86.35 ± 3.18 | 79.56 ± 2.97 | 79.00 ± 3.08 | 83.33 ± 4.56 |

### *2.4. Implementation and training*

The 10 × 10-fold cross validation strategy was used to evaluate the experimental performance. In each 10-fold cross-validation process, the raw dataset is first divided into ten portions, with nine used as the training set and the remaining one as the testing set. This procedure was repeated ten times. The processed dataset contains 128 normal unilateral kidneys and 140 abnormal unilateral kidneys. During the training process, data augmentation was only applied to the images in the training set.

During the training process of the four models, the same training configuration was applied. All models were implemented and trained on PyTorch 1.11 with a NVIDIA 3090TI GPU(24 GB), using the Adam optimizer with an initial learning rate of 0.0001. The learning rate decreases by a factor of 0.1 every 20 epochs. All model training batch sizes were set to 32, and the number of iterations was set to 100 epochs. We used Cross Entropy as the loss function for training the model.

### *2.5. Evaluation metrics*

Six evaluation metrics are used to assess the final performance of the model: Accuracy, Area Under the Curve (AUC), Negative Predictive Value (NPV), Positive Predictive Value (PPV), Sensitivity, and Specificity. Some evaluation metrics are represented in Equations (1)-(5) below.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$NPV = \frac{TN}{TN + FN} \tag{2}$$

$$PPV = \frac{TP}{TP + FP} \tag{3}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{4}$$

$$Specificity = \frac{TN}{TN + FP} \tag{5}$$

where TP, FP, FN, and TN denote the number of true positives, false positives, false negatives, and true negatives, respectively. More specifically, in this paper, positive samples represent the cases that require preventive antibiotic treatments, while negative samples represent the cases that do not require preventive antibiotic treatments.

## 3. Experiment and result

First, we compared the classification results of the four classic models using a 10 × 10-fold cross-validation strategy. Next, we conducted further experiments with AlexNet, the best of the four models, to evaluate our DL-assisted diagnosis strategy for children with febrile pyelonephritis.

### *3.1. The experimental results of the four models*

Table 1 lists the classification results of the four models. Among the four models, AlexNet achieves the best performance across six indicators, with the accuracy of 84.05% and the AUC of 84.21%. Compared to AlexNet, deeper models like VGG16, ResNet50, and DenseNet201 trained on the same data achieve a relatively low performance. These three models with larger parameters are more prone to overfitting when the training data is insufficient. In this experiment, ResNet50 had the lowest accuracy of 75.43%, which was 8.62% lower than AlexNet. In contrast, the deeper model DenseNet201 achieved an accuracy of 81.20%, which is only 2.85% lower than that of AlexNet. Based on the experimental results, it is demonstrated that the dense connections in DenseNet can capture low-dimensional information, have a regularizing effect, and enable the model to perform well even with limited data.

### *3.2. Extra experiments on AlexNet*

To verify whether using the images including single kidney of subject as model input would result in superior outcomes compared to using the images including two kidneys of subject as input, we conducted extra experiments using image including two kidneys

**Table 2**
The comparison between AlexNet using images that include two kidneys as input and AlexNet using single kidney images as input. * and ** indicate the statistically significant differences of $p \leq 0.05$ and $p \leq 0.01$ by the paired *t*-test, respectively.

| Method | Metric | | | | | |
|---|---|---|---|---|---|---|
| | Accuracy (%) | AUC | PPV (%) | NPV (%) | Sensitivity (%) | Specificity (%) |
| AlexNet (single kidney) | 84.05 ± 0.95 | 84.21 ± 0.93 | 87.79 ± 1.81 | 81.90 ± 1.45 | 81.71 ± 2.10 | 86.70 ± 2.27 |
| AlexNet (two kidneys) | 80.92 ± 2.73** | 79.09 ± 3.16** | 86.34 ± 2.57 | 75.10 ± 4.37** | 84.29 ± 2.67** | 74.93 ± 4.98** |
| AlexNet w/o data augmentation (single kidney) | 81.87 ± 1.34** | 82.06 ± 1.40** | 86.22 ± 1.79 | 79.41 ± 1.08** | 78.86 ± 1.32** | 85.27 ± 2.26 |
| AlexNet w/o data augmentation (two kidneys) | 73.26 ± 1.61** | 72.34 ± 1.71** | 82.45 ± 2.25** | 63.75 ± 3.51** | 75.55 ± 3.19** | 69.13 ± 4.45** |

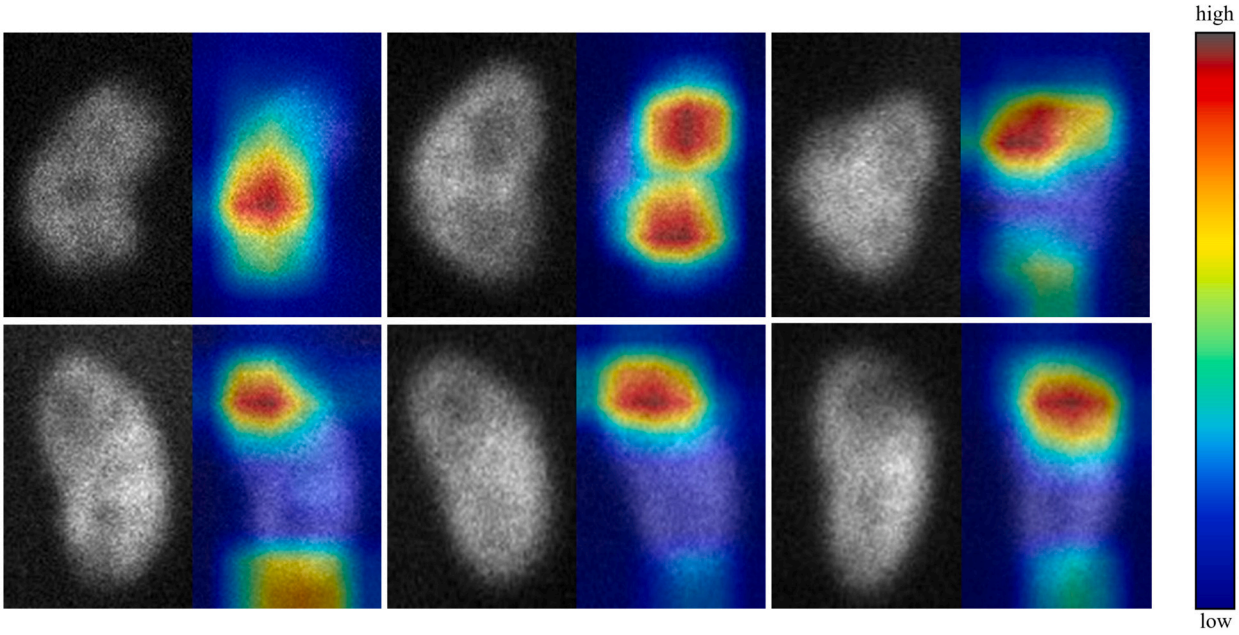



**Fig. 4.** Visualize the heat maps of AlexNet for diseased kidneys using Grad-CAM. The blue area and the red area represent lower and higher contribution scores to the classification results of AlexNet, respectively.

as the input for the AlexNet model. The results are presented in Table 2. It can be found that without data augmentation, using the image including two kidneys resulted in accuracy decrease by 8.61% compared to using a single kidney. By using data augmentation, the accuracy of images including two kidneys obtained through AlexNet increased from 73.26% to 80.92%. Nevertheless, due to the imbalance in sample size between the two categories, the images of two kidneys exhibited poor performance in NPV and specificity, measuring at 75.10% and 74.93%, respectively. Even with the same data augmentation, the accuracy and AUC of images of a single kidney were 3.13% and 5.11% higher than those of images including two kidneys, respectively. This implies that using images of a single kidney can make the network focus more easily on important region information. Moreover, as shown in Table 2, regardless of whether images including single kidney or two kidneys were used, the accuracies of the AlexNet model were improved after data augmentation. This proves the effectiveness of the data augmentation methods used.

Furthermore, in order to verify the reliability of the results obtained by AlexNet, Grad-CAM [29] was used to visualize the regions of interest of the model. As shown in Fig. 4, we visualized the heat maps of the output features from the last layer of AlexNet. Regardless of whether the input is the left kidney or the right kidney, the AlexNet model exhibits a remarkable ability to discern the precise position and intricate shape of the kidneys. It can be observed that AlexNet model focuses more attention on the kidney region rather than the noise in the background. After analyzing the key areas, we found that the areas of focus closely aligned with those identified by human doctors based on their accumulated experience, indicating that AlexNet model could serve as a baseline and effective method to classify the two groups of children.

In addition, to facilitate the application of DL model for computer-assisted decision support for the diagnosis and treatment of febrile pyelonephritis, we developed a software program interface as shown in Fig. 5. The system includes three modules, i.e., model selection, results, and image display. In the Model selection module, the user can select different pre-trained models to analyze the loaded images. In the Results module, the classification label and reliability are provided. In the module of image display, the loaded image and the heatmap of the selected model are displayed. In the current version, the above-mentioned four models were integrated for selection, any new models can be added into the system in future. Using the visual interface, doctors can comprehensively evaluate the results produced by different models.

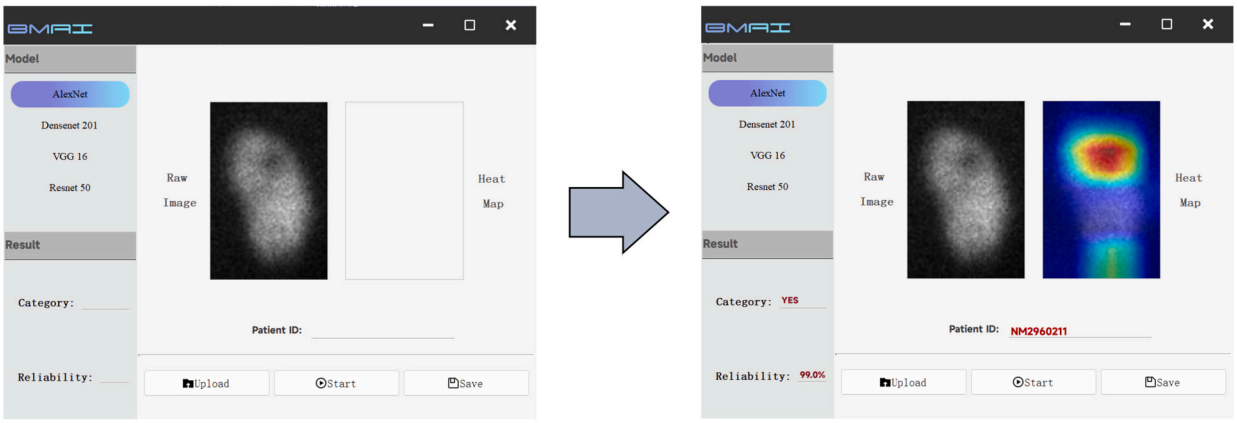



**Fig. 5.** Illustration of software interface for computer-assisted decision support for the usage of preventive antibacterial therapy in children with febrile pyelonephritis. The left subplot shows the status of the image loaded into the software, and the right subplot shows the results obtained from the DL model, including classification results, reliability, patient ID, and corresponding heatmap.

## 4. Discussion and conclusion

Childhood urinary tract infections pose a significant health burden to society. Data from the United States show that newly diagnosed urinary tract infections in children account for 2.4%-2.8% of all children each year, and the cost of hospitalization costs alone exceed $ 18 billion [30]. The diagnosis of urinary tract infections is mainly based on clinical manifestations and laboratory tests, but the general clinical and laboratory indicators are not very specific, especially for diagnosing pyelonephritis in infants and young children, which is challenging and often leads to misdiagnosis due to its lack of special clinical manifestations. The clinical value of $^{99m}$Tc – DMSA is derived from its comparison with intravenous urography and ultrasound. It exhibits a higher sensitivity in diagnosing acute pyelonephritis and predicting prognosis compared to the latter two imaging techniques. Children with urinary tract infections not only endure the clinical symptoms during infection but also face the risk of long-term consequences. Therefore, $^{99m}$Tc – DMSA renal static scan is very important for early diagnosis of pyelonephritis and rational treatment, which will reduce the huge burden of treatment costs for childhood urinary tract infections to society and families [31]. The traditional diagnosis of pyelonephritis relies on visual inspection of $^{99m}$Tc – DMSA images by physicians, and due to the presence of noise and subjectivity, the diagnostic opinions of different physicians may not be consistent.

In recent years, the deep learning-based methods are widely used for computer-aided diagnosis, and have been used for the prediction of recurrence of urinary tract infection and the denoising of 99mTC-DMSA images [13]. However, no related studies have investigated whether the children first diagnosed with febrile pyelonephritis require preventive antibiotic treatments. In this study, for the first time, we checked the effectiveness of deep learning models for this task, aiming to provide auxiliary guidance for the diagnosis of pyelonephritis. To achieve this, we employed four different convolutional neural networks to classify $^{99m}$Tc – DMSA images. To minimize image noise and expand the limited dataset, we used a series of methods to ensure that the model receives optimal input. Based on the experiments, we discovered that AlexNet outperformed the other three models in six evaluation metrics. For AlexNet, on the one hand, we verified the effectiveness of the techniques employed through comparative experiments. On the other hand, we used Grad-CAM technique to conduct visual analysis of the AlexNet model to ensure its reliability. As a final step, we presented the specific evaluation results of the four models used through the interface that can be used to provide computer-assisted decision support for the diagnosis of febrile pyelonephritis.

Several limitations should be mentioned in this study. Although AlexNet has shown promising performance in the task of classifying $^{99m}$Tc – DMSA images, the current accuracy has not yet reached an ideal level to directly use in the clinical scenario. Meanwhile, AlexNet may not show satisfactory performance for all cases of pyelonephritis. $^{99m}$Tc – DMSA imaging cannot differentiate between image abnormalities caused by bacterial infections and those caused by other factors, such as fungal infections. A small subset of inflammation resulting from non-bacterial infections requires a comprehensive medical history and additional tests for proper differentiation. In addition, we did not conduct more experiments with more advanced models in this preliminary study. High-performance models need to be developed in future studies, and some other technologies, such as transfer learning [32], can be adopted to enhance the classification performance with limited numbers of $^{99m}$Tc – DMSA images.

The sample size in the current dataset is limited, therefore, to enhance the robustness and generalizability of this study, it is imperative to gather extensive datasets for future investigations. The future datasets should include diverse cohort populations, and be collected from multiple clinical centers [33]. These diverse samples will be beneficial to develop robust and generalizable DL models. In this study, only the $^{99m}$Tc – DMSA images were adopted to implement the diagnosis on the febrile pyelonephritis. However, emerging evidence suggests that the fusion of multimodal medical images could potentially enhance the accuracy of computer-aided diagnosis [34], such as medical diagnostics and segmentation [35,36]. In our future work, we will combine the multimodal medical images and design fusion algorithm to improve the diagnostic performance for febrile pyelonephritis.

In conclusion, this preliminary study contributes valuable insights into determining the need for preventive antibiotic usage in children first diagnosed with febrile pyelonephritis with the deep learning technology, offering a potential solution to the clinical

aided diagnosis of febrile pyelonephritis. In the future, more efforts should be focused on developing advanced methods and acquiring a large dataset to boost the diagnosis performance.

## CRediT authorship contribution statement

**Zhengguo Chen:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Ning Li:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology. **Zhu Chen:** Writing – review & editing, Conceptualization. **Li Zhou:** Writing – review & editing, Conceptualization. **Liming Xiao:** Writing – review & editing, Conceptualization. **Yangsong Zhang:** Writing – review & editing, Supervision, Methodology, Conceptualization.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The data that has been used is confidential.

## Acknowledgements

The authors thank the reviewers and editors for their insightful comments. This work was supported in part by the National Natural Science Foundation of China under Grant No. 62076209, and in part by the NHC Key Laboratory of Nuclear Technology Medical Transformation (MIANYANG CENTRAL HOSPITAL) under Grant No. 2021HYX026 and No. 2022HYX012.

## References

[1] Lindsey Korbel, Marianella Howell, John David Spencer, The clinical diagnosis and management of urinary tract infections in children and adolescents, Paediatr. Int. Child Health 37 (4) (2017) 273–279.
[2] Chen Lei, Mark Douglas Baker, Racial and ethnic differences in the rates of urinary tract infections in febrile infants in the emergency department, Pediatr. Emerg. Care 22 (7) (2006) 485–487.
[3] Kathryn O'Brien, Naomi Stanton, Adrian Edwards, Kerenza Hood, Christopher C. Butler, Prevalence of urinary tract infection (UTI) in sequential acutely unwell children presenting in primary care: exploratory study, Scand. J. Prim. Health Care 29 (1) (2011) 19–22.
[4] Nader Shaikh, Mary Ann Haralam, Marcia Kurs-Lasky, Alejandro Hoberman, Association of renal scarring with number of febrile urinary tract infections in children, JAMA Pediatr. 173 (10) (2019) 949–952.
[5] Wei Yang, Qinghan Jiao, Haiyan Wang, Weizhen Chen, Hongxiang Yao, Is technetium-99m dimercaptosuccinic acid renal scintigraphy available for predicting vesicoureteral reflux in children with first febrile urinary tract infection under the age of 24 months?, Nucl. Med. Commun. 43 (11) (2022) 1128–1135.
[6] Tej K. Mattoo, Nader Shaikh, Caleb P. Nelson, Contemporary management of urinary tract infection in children, Pediatrics 147 (2) (2021).
[7] Stephen S. Yang, Jeng-Daw Tsai, Akihiro Kanematsu, Chang-Hee Han, Asian guidelines for urinary tract infection in children, J. Infect. Chemother. 27 (11) (2021) 1543–1554.
[8] Banafsheh Arad, Abolfazl Mahyar, Mahmoud Vandaie, Sonia Oveisi, Prediction of vesicoureteral reflux by ultrasonography and renal scan in children, Glob. Pediatr. Health 9 (2022) 2333794X221107826.
[9] Nisha Jacob, Seshagiri Koripadu, Harishchandra Yanamandala, A study to determine risk factors for renal scarring as detected by dimercaptosuccinic acid scan in children with urinary tract infection, Int. J. Contemp. Pediatr. 8 (2021) 1333.
[10] Kenji Hirata, Hiroyuki Sugimori, Noriyuki Fujima, Takuya Toyonaga, Kohsuke Kudo, Artificial intelligence for nuclear medicine in oncology, Ann. Nucl. Med. (2022) 1–10.
[11] Felix Nensa, Aydin Demircioglu, Christoph Rischpler, Artificial intelligence in nuclear medicine, J. Nucl. Med. 60 (Supplement 2) (2019) 29S–37S.
[12] Edward H. Herskovits, Artificial intelligence in molecular imaging, Ann. Transl. Med. 9 (9) (2021).
[13] Jagrati Chaudhary, Ankita Phulia, Anil Kumar Pandey, Param D. Sharma, Chetan Patel, Denoising Tc-99m DMSA images using denoising convolutional neural network with comparison to a block matching filter, Nucl. Med. Commun. 44 (8) (2023) 682–690.
[14] Hyunjong Lee, Beongwoo Yoo, Minki Baek, Joon Young Choi, Prediction of recurrent urinary tract infection in paediatric patients by deep learning analysis of 99mTc-DMSA renal scan, Diagnostics 12 (2) (2022) 424.
[15] Shannon E. O'Reilly, Donika Plyku, George Sgouros, Frederic H. Fahey, S. Ted Treves, Eric C. Frey, Wesley E. Bolch, A risk index for pediatric patients undergoing diagnostic imaging with 99mTc-dimercaptosuccinic acid that accounts for body habitus, Phys. Med. Biol. 61 (6) (2016) 2319.
[16] Selçuk Varol, Faruk Öktem, Abdurrahim Koçyiğit, Ayşegül Doğan Demir, Ersin Karataş, Mehmet Aydın, Nilüfer Göknar, Tümay İpekçi, The impact of Technetium-99m dimercapto-succinic acid scintigraphy on DNA damage and oxidative stress in children, Int. J. Clin. Pract. 75 (11) (2021) e14810.
[17] N.A. Nikolov, D.A. Supruniuk, A.N. Smetaniuk, Stochastic properties of nephroscintigraphic images with 99mTc-DMSA, in: 2013 IEEE XXXIII International Scientific Conference Electronics and Nanotechnology (ELNANO), 2013, pp. 246–250.
[18] Justin Salamon, Juan Pablo Bello, Deep convolutional neural networks and data augmentation for environmental sound classification, IEEE Signal Process. Lett. 24 (3) (2017) 279–283.
[19] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, Ruslan Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (1) (2014) 1929–1958.
[20] Connor Shorten, Taghi M. Khoshgoftaar, A survey on image data augmentation for deep learning, J. Big Data 6 (1) (2019) 1–48.
[21] Agnieszka Mikołajczyk, Michał Grochowski, Data augmentation for improving deep learning in image classification problem, in: 2018 International Interdisciplinary PhD Workshop (IIPhDW), 2018, pp. 117–122.
[22] Wei-Chung Shia, Li-Sheng Lin, Dar-Ren Chen, Classification of malignant tumours in breast ultrasound using unsupervised machine learning approaches, Sci. Rep. 11 (1) (2021) 1–11.

[23] Niall O'Mahony, Sean Campbell, Anderson Carvalho, Suman Harapanahalli, Gustavo Velasco Hernandez, Lenka Krpalkova, Daniel Riordan, Joseph Walsh, Deep learning vs. traditional computer vision, in: Kohei Arai, Supriya Kapoor (Eds.), Advances in Computer Vision, Springer International Publishing, Cham, 2020, pp. 128–144.
[24] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, Imagenet classification with deep convolutional neural networks, Commun. ACM 60 (6) (2017) 84–90.
[25] Karen Simonyan, Andrew Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint, arXiv:1409.1556, 2014.
[26] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.
[27] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, Kilian Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 4700–4708.
[28] Xavier Glorot, Antoine Bordes, Yoshua Bengio, Deep sparse rectifier neural networks, in: Geoffrey Gordon, David Dunson, Miroslav Dudík (Eds.), Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, Fort Lauderdale, FL, USA, 11–13 Apr 2011, in: Proceedings of Machine Learning Research, vol. 15, PMLR, pp. 315–323,.
[29] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, Dhruv Batra, Grad-CAM: visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 618–626.
[30] Andrew L. Freedman, Urologic Diseases in America Project, Urologic diseases in North America Project: trends in resource utilization for urinary tract infections in children, J. Urol. 173 (3) (2005) 949–954.
[31] Lisette A't Hoen, Guy Bogaert, Christian Radmayr, Hasan S. Dogan, Rien J.M. Nijman, Josine Quaedackers, Yazan F. Rawashdeh, Mesrur S. Silay, Serdar Tekgul, Nikita R. Bhatt, et al., Update of the EAU/ESPU guidelines on urinary tract infections in children, J. Pediatr. Urol. 17 (2) (2021) 200–207.
[32] Sinno Jialin Pan, Qiang Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (10) (2010) 1345–1359.
[33] Debesh Jha, Ashish Rauniyar, Abhiskek Srivastava, Desta Haileselassie Hagos, Nikhil Kumar Tomar, Vanshali Sharma, Elif Keles, Zheyuan Zhang, Ugur Demir, Ahmet Topcu, Anis Yazidi, Jan Erik Håakegård, Ulas Bagci, Ensuring trustworthy medical artificial intelligence through ethical and philosophical principles, arXiv preprint, arXiv:1409.1556, 2023.
[34] Muhammad Adeel Azam, Khan Bahadar Khan, Sana Salahuddin, Eid Rehman, Sajid Ali Khan, Muhammad Attique Khan, Seifedine Kadry, Amir H. Gandomi, A review on multimodal medical image fusion: compendious analysis of medical modalities, multimodal databases, fusion techniques and quality metrics, Comput. Biol. Med. 144 (2022) 105253.
[35] Takahiro Hosokawa, Mayuki Uchiyama, Yutaka Tanami, Yumiko Sato, Yasuharu Wakabayashi, Eiji Oguma, Incidence of renal scarring on technetium-99 m dimercaptosuccinic acid renal scintigraphy after acute pyelonephritis, acute focal bacterial nephritis, and renal abscess, Ann. Nucl. Med. 37 (3) (March 2023) 176–188.
[36] Leonardo Rundo, Alessandro Stefano, Carmelo Militello, Giorgio Russo, Maria Gabriella Sabini, Corrado D'Arrigo, Francesco Marletta, Massimo Ippolito, Giancarlo Mauri, Salvatore Vitabile, Maria Carla Gilardi, A fully automatic approach for multimodal PET and MR image segmentation in gamma knife treatment planning, Comput. Methods Programs Biomed. 144 (2017) 77–96.