

## Article

# Deep Learning Model for Computer-Aided Diagnosis of Urolithiasis Detection from Kidney–Ureter–Bladder Images

Yi-Yang Liu <sup>1,2</sup>, Zih-Hao Huang <sup>1</sup>  and Ko-Wei Huang <sup>1,\*</sup> 

<sup>1</sup> Department of Electrical Engineering, National Kaohsiung University of Science and Technology, Kaohsiung City 80778, Taiwan

<sup>2</sup> Department of Urology, Kaohsiung Chang Gung Memorial Hospital and Chang Gung University College of Medicine, Kaohsiung City 83301, Taiwan

\* Correspondence: elone.huang@nkust.edu.tw

**Abstract:** Kidney–ureter–bladder (KUB) imaging is a radiological examination with a low cost, low radiation, and convenience. Although emergency room clinicians can arrange KUB images easily as a first-line examination for patients with suspicious urolithiasis, interpreting the KUB images correctly is difficult for inexperienced clinicians. Obtaining a formal radiology report immediately after a KUB imaging examination can also be challenging. Recently, artificial-intelligence-based computer-aided diagnosis (CAD) systems have been developed to help clinicians who are not experts make correct diagnoses for further treatment more effectively. Therefore, in this study, we proposed a CAD system for KUB imaging based on a deep learning model designed to help first-line emergency room clinicians diagnose urolithiasis accurately. A total of 355 KUB images were retrospectively collected from 104 patients who were diagnosed with urolithiasis at Kaohsiung Chang Gung Memorial Hospital. Then, we trained a deep learning model with a ResNet architecture to classify KUB images in terms of the presence or absence of kidney stones with this dataset of pre-processed images. Finally, we tuned the parameters and tested the model experimentally. The results show that the accuracy, sensitivity, specificity, and F1-measure of the model were 0.977, 0.953, 1, and 0.976 on the validation set and 0.982, 0.964, 1, and 0.982 on the testing set, respectively. Moreover, the results demonstrate that the proposed model performed well compared to the existing CNN-based methods and was able to detect urolithiasis in KUB images successfully. We expect the proposed approach to help emergency room clinicians make accurate diagnoses and reduce unnecessary radiation exposure from computed tomography (CT) scans, along with the associated medical costs.

**Keywords:** computer-aided diagnosis; kidney–ureter–bladder; kidney stone; deep learning; residual network



**Citation:** Liu, Y.-Y.; Huang, Z.-H.; Huang, K.-W. Deep Learning Model for Computer-Aided Diagnosis of Urolithiasis Detection from Kidney–Ureter–Bladder Images. *Bioengineering* **2022**, *9*, 811. <https://doi.org/10.3390/bioengineering9120811>

Academic Editors: Faezeh Marzbanrad and Chiranjibi Sitaula

Received: 10 November 2022

Accepted: 12 December 2022

Published: 16 December 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Studies based on data from seven countries (Italy, Germany, Scotland, Spain, Sweden, Japan, and the United States) have shown that the prevalence and incidence of kidney stones have been increasing globally [1–3]. To address this problem, various methods have been developed to detect and treat kidney stones [4,5]. Computed tomography (CT) is a particularly accurate diagnostic method, with sensitivity and specificity ranging from 94% to 100% and 92% to 94.2% for kidney stones, respectively [6,7]. Therefore, CT is the gold standard for kidney stone diagnosis. However, CT is costly and requires a higher radiation dose than plain film X-ray imaging. For example, the radiation dose of an abdominal CT scan ranges from 8 to 34 mGy [8,9], in contrast to the lower dose of 2.47 mGy required to record a kidney–ureter–bladder (KUB) image [10]. Similarly, a stomach CT requires 50 times the radiation dose of a plain film stomach X-ray [11]. Although low-dose CT reduces the radiation dose from 25 to 17 mGy for abdominal CT scans, this value is still higher than that of plain film X-ray imaging [12]. Therefore, plain film X-ray imaging may be considered as a cost-effective alternative to CT, which also causes less harm to the

body. However, X-ray images also have a propensity for false positives owing to their 2D nature, and their resolution does not suffice to identify abnormalities in dense tissues [13]. For example, KUB images have sensitivities of only 44–77% and specificities of 80–87% in kidney stone detection [14], which are considerably inferior values compared to those of CT. Improving the sensitivity of KUB imaging for use with kidney stones could therefore allow X-ray scans to be used as a widely applicable option to diagnose the condition, which would also reduce medical costs.

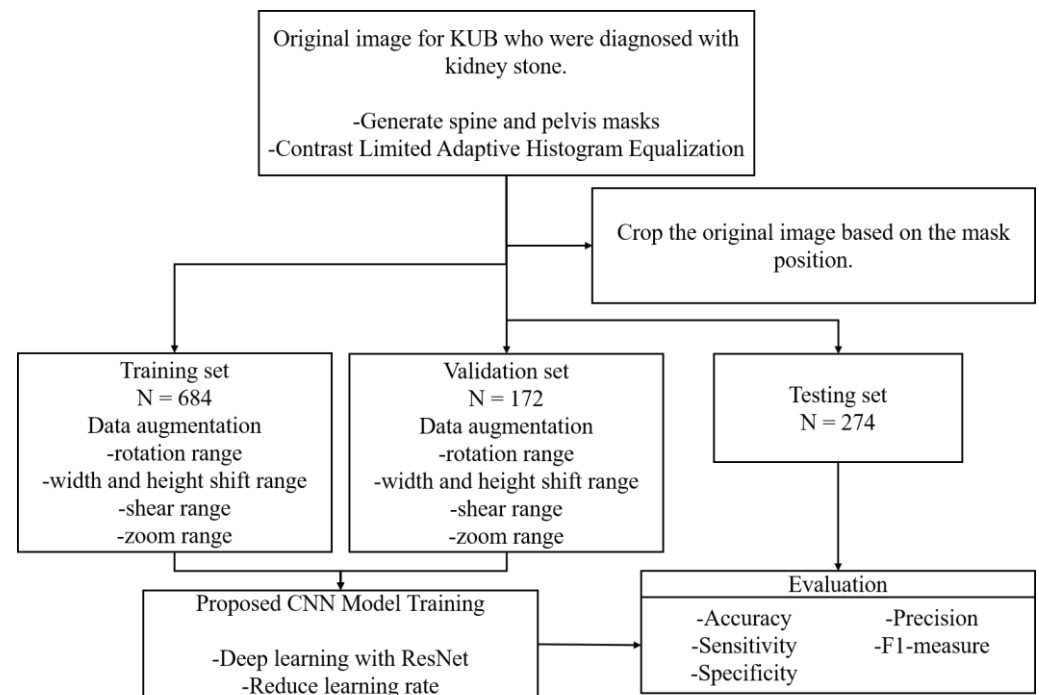
Medical imaging has become increasingly important in clinical diagnosis. X-rays, magnetic resonance imaging (MRI), and CT are among the most common medical imaging modalities. Medical image processing often depends on the experience of radiologists, who must analyze these images and draw conclusions using a subjective approach. As the variety and number of medical imagery techniques have significantly increased in recent years, manual analysis has become increasingly time-consuming and labor-intensive. To address this problem, machine learning models have been used to replicate human visual perception mechanisms to enable computational systems to automatically classify medical images as diagnostic aids. Computational techniques have become significantly more powerful over the last several decades owing to rapid advances in AI and computing hardware, and the use of computer-aided methods to analyze and process medical imagery has become incredibly useful for diagnosticians both in classifying and in augmenting those images. Thus, computer-aided diagnosis (CAD) has become theoretically and practically significant as an important trend in medical science. The use of computer vision to automatically analyze and process medical images has several unique advantages [15–17]. For example, this approach leverages the immense computing power of modern hardware to achieve rapid and accurate analysis and processing, which renders its findings immune to fatigue or cognitive issues with information overload. Furthermore, computer technologies and networks can enable the rapid transfer of clinical data to facilitate the rapid and accurate diagnosis of patients in remote locations. Machine learning algorithms designed to diagnose and detect various medical conditions have become a topic of active research, and the accuracy of artificial intelligence classifiers used to predict various related data of patients with kidney stones has also increased [18]. Since the emergence of the first convolutional neural network (CNN) (LeNet-5 [19]) in 1989, CNN models have continued to improve, and deep CNNs have been shown to perform extremely well in medical image processing [20–23]. The accuracy of CAD methods has also benefited from the progressive improvement of such models [24,25]. In urology, several studies have considered the use of neural networks to aid in the diagnosis of urinary diseases based on CT imaging [26–28]. The application of CAD to X-ray examinations has also yielded impressive results. For example, a CNN model trained to diagnose urinary tract stones from plain film X-ray imaging using pre-processed images showed a sensitivity of 89.6% and PPV of 56.9% for kidney stones [29].

Thus far, KUB imaging is still considered as a first-line examination for urolithiasis detection in the emergency room due to its convenience, low cost, and low radiation dose. However, only highly experienced urologists or radiologists can diagnose urolithiasis correctly from KUB images. Furthermore, emergency physicians who arrange KUB images cannot immediately obtain a formal report from the experts. Hence, emergency physicians without the necessary specialized experience are highly likely to either make incorrect diagnoses or choose to arrange non-contrast CT scans for such patients, which may delay further treatment or increase medical costs and the radiation dose. In this study, to address this challenge, we constructed a CAD system based on a deep learning model trained to help emergency physicians make correct diagnoses of urolithiasis from KUB images.

## 2. Materials and Methods

### 2.1. Datasets

The protocol of the present study was approved by the Institutional Review Board of Kaohsiung Chang Gung Memorial Hospital. A total of 355 KUB images were retrospectively collected from 104 patients from Kaohsiung Chang Gung Memorial Hospital who were diagnosed with stones in their upper urinary tract. The presence of stones in the upper urinary tract shown in these 355 images was formally reported by radiologists and then confirmed on a case-by-case basis by two experienced urologists specializing in urolithiasis. The set of KUB images was first divided into groups of training images with single or multiple urinary tract stones, and the dataset was augmented through various image pre-processing operations to produce a total of 1130 images. Then, these 1130 images were divided into three datasets, with 856 images used to train the network, with 80% (684 images) allocated to the training process itself and 20% (172 images) for validation. The remaining 274 images were used to evaluate the performance of the trained model based on several metrics and to test its generalizability. A flowchart of the work performed in this study is shown in Figure 1.



**Figure 1.** Process flow used in this study.

### 2.2. Image Pre-Processing

First, a Mask R-CNN model was trained to detect the spine and pelvis bones in the KUB images [30,31], and the trained model was applied to mask most of the high-brightness regions. The images were then centered on the spine, and the area above the pelvis region was segmented. Because identifying abnormalities in highly dense tissues using plain film X-ray images is difficult, we aimed to exclude factors that tend to lead to the misidentification of features around the kidneys [13]. Furthermore, because of the characteristics of plain film X-ray imaging, dense tissues appear with higher brightness. Hence, histogram equalization [32] can easily lead to overexposure of the image and thus influence the detection of urinary tract stones. The effects of masking on an X-ray image may be observed from a histogram. Contrast-limited adaptive histogram equalization (CLAHE) has been used to enhance contrast in KUB imaging [33], which allows stones to be distinguished from the background through their brightness and also prevents overexposure from excessively high brightness. In this study, we compared the effects of

histogram equalization and CLAHE on the KUB images. Finally, patches with a size of  $100 \times 100$  pixels were cropped from the pre-processed X-ray plain films. Patches containing a stone were cropped with the stone at the center, whereas patches without a stone were randomly cropped from the pre-processed plain film X-ray images [29].

### 2.3. Data Augmentation

Numerous studies have shown that data augmentation is effective in preventing overfitting, which is more likely to occur for CNN models trained with smaller datasets [34–37]. In particular, relatively large datasets of medical images for analysis are often difficult to obtain. Furthermore, the generalizability of learning models depends on the diversity of the data samples [38–40]; the more generalizable a model is, the more accurate its results with images that were not present in the training dataset. During training, 100% accuracy can be achieved very quickly, although the prediction accuracy of a model trained in this way is typically reduced. To prevent overfitting and improve sample diversity, we performed data augmentation prior to the training process by rotating, vertically and horizontally translating, magnifying/shrinking, and shear-mapping the original images. However, in contrast to conventional data augmentation methods, the images in the training set were randomly augmented after each iteration of the training process to produce a dynamically augmented dataset. This approach also greatly reduces memory consumption.

### 2.4. Deep Learning Models

We adopted a ResNet-50 architecture as the CNN model in this study. Many studies have shown that the fineness of detail that can be extracted by a CNN increases with the depth of the network. However, He et al. (2016) demonstrated that performance degrades if the depth increases beyond a certain point [41]. Residual network (ResNet) architectures are based on residual blocks comprising convolutional, activation, and batch normalization (BN) layers  $F(x)$ , and a shortcut connection that reproduces the input  $x$ . Because the output of a residual block is  $H(x) = F(x) + x$ , the layers in a traditional network effectively learn the difference between the true output and  $x$ , i.e., the residual, as shown in Figure 2. Therefore, for the simple case in which the network has not learned any features and the input is already optimal,  $F(x)$  is approximately 0, or  $H(x) = x$  (i.e., the identity relation). This solves the degradation problem and allows for extremely deep networks. The ResNet architecture is shown in Figure 3. By employing deep learning models for image classification, images can be automatically classified and labeled for various applications [42].

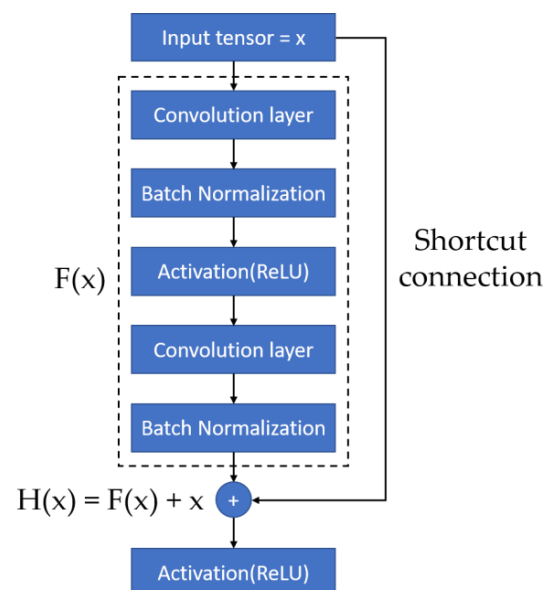
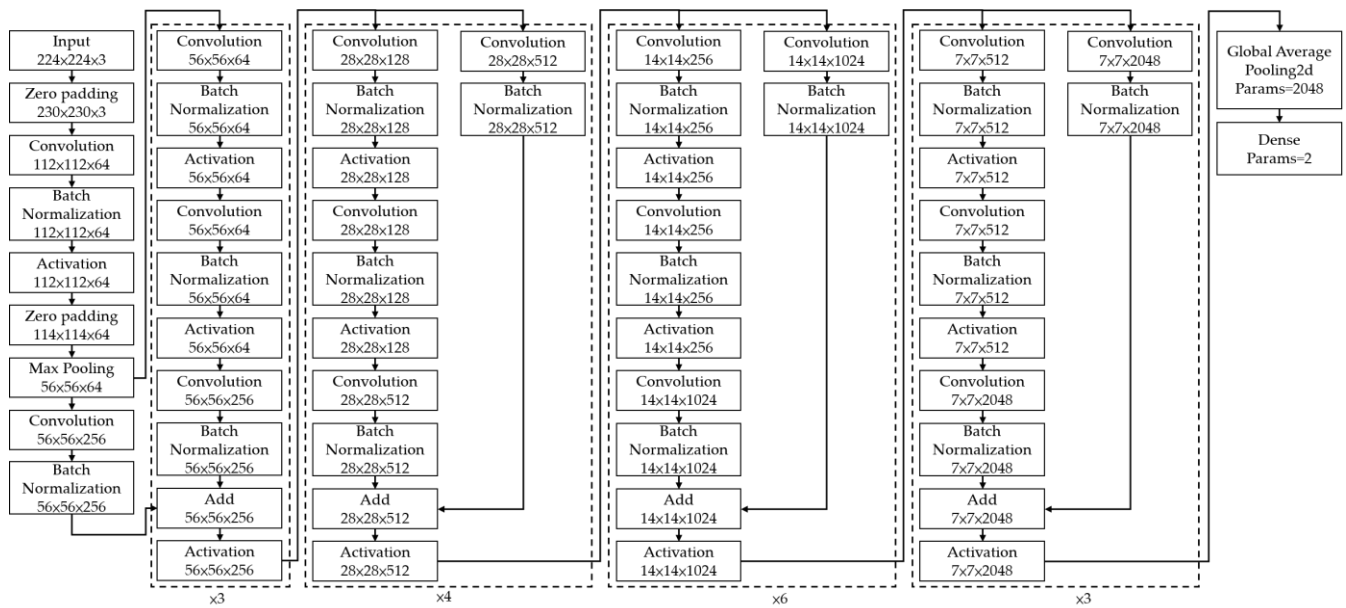


Figure 2. Residual Block [41].



**Figure 3.** Residual Network Architecture [41].

## 2.5. Technical Details and Evaluation Metrics

The data were divided into training, validation, and testing sets. The validation set was taken from the training set at a 20:80 ratio. The testing set consisted of 24% of the total data and was used to evaluate the accuracy of the trained ResNet model. The inputs of the image classification model consisted of images with a size of  $224 \times 224$  px, and the urinary tract stone images were diversified using data augmentation techniques (random rotation, horizontal/vertical translation, magnification/zooming out, and shear-mapping). The Keras API with the Tensorflow platform (version 2.9.1) was used to construct the ResNet model. Ranger [43], which was created by combining RAdam [44] with LookAhead [45], was used as an optimizer. As with Adam, RAdam converges quickly and achieves a level of optimality similar to that of SGD. Furthermore, RAdam converges similarly with different learning rates, whereas Adam and SGD are much more sensitive to the learning rate and require optimization. Binary cross-entropy was used as a loss function. The predictions were used to construct a confusion matrix of four possible outcomes (see Figure 4). Correct predictions are either true positive (TP) or true negative (TN), whereas incorrect predictions are either false positive (FP) or false negative (FN). These outcomes were used to construct seven metrics to evaluate the performance of the model, including accuracy, sensitivity, specificity, precision, F1-measure, the receiver operator characteristic (ROC) curve, and the area under the ROC curve (AUC). Accuracy is defined as

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

Although this is a simple metric, accuracy is susceptible to bias for unbalanced training data. Therefore, we also used the four other metrics mentioned above. Sensitivity, also known as recall, provides the proportion of patients with kidney stones who were correctly predicted as having the condition. Sensitivity is given by

$$\text{sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

Specificity provides the proportion of patients without kidney stones who were correctly predicted as negative for the condition, and is given by

$$\text{specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}} \quad (3)$$

Precision is the proportion of patients who actually had kidney stones among all persons predicted to have the condition, and is given by

$$\text{precision} = \frac{TP}{TP + FP} \quad (4)$$

The F-measure is a comprehensive measure of performance, of which F1-measure is a special case. If  $\beta$  is equal to 1, the F1-measure will equally reflect both recall and precision. However, if  $\beta$  is greater than 1, the F1-measure depends more on the recall than the precision, and vice versa. The F1-measure is given by Equation (5); higher values indicate better performance.

$$F_{\beta} - \text{measure} = \left(1 + \beta^2\right) \frac{\text{Precision} \times \text{Recall}}{(\beta^2 \times \text{Precision}) + \text{Recall}} \quad (5)$$

The last two metrics are the ROC curve and AUC. The ROC curve is obtained by plotting the true positive rate (TPR) (y-axis) against the false positive rate (FPR) (x-axis). The TPR is the proportion of positive predictions that are actually positive, and the FPR is the proportion of positive predictions that are actually negative. Therefore, the ROC curve represents the relationship between the FPR and TPR in the model. Because the (0, 1) point corresponds to a perfect classification, the performance of a model is proportional to the closeness of its ROC curve to the top-left corner. Similarly, the AUC is the area under the ROC curve, which increases as the ROC curve approaches the top-left corner. Hence, the performance of a classification model is directly proportional to the AUC.

	Positive	Negative	
Positive	True Positive(TP)	False Negative(FN)	<b>Sensitivity</b> $\frac{TP}{TP + FN}$
Negative	False Positive(FP)	True Negative(TN)	<b>Specificity</b> $\frac{TN}{TN + FP}$
	<b>Precision</b> $\frac{TP}{TP + FP}$	<b>Negative Predictive Value(NPV)</b> $\frac{TN}{TN + FN}$	<b>Accuracy</b> $\frac{TP + TN}{TP + TN + FP + FN}$

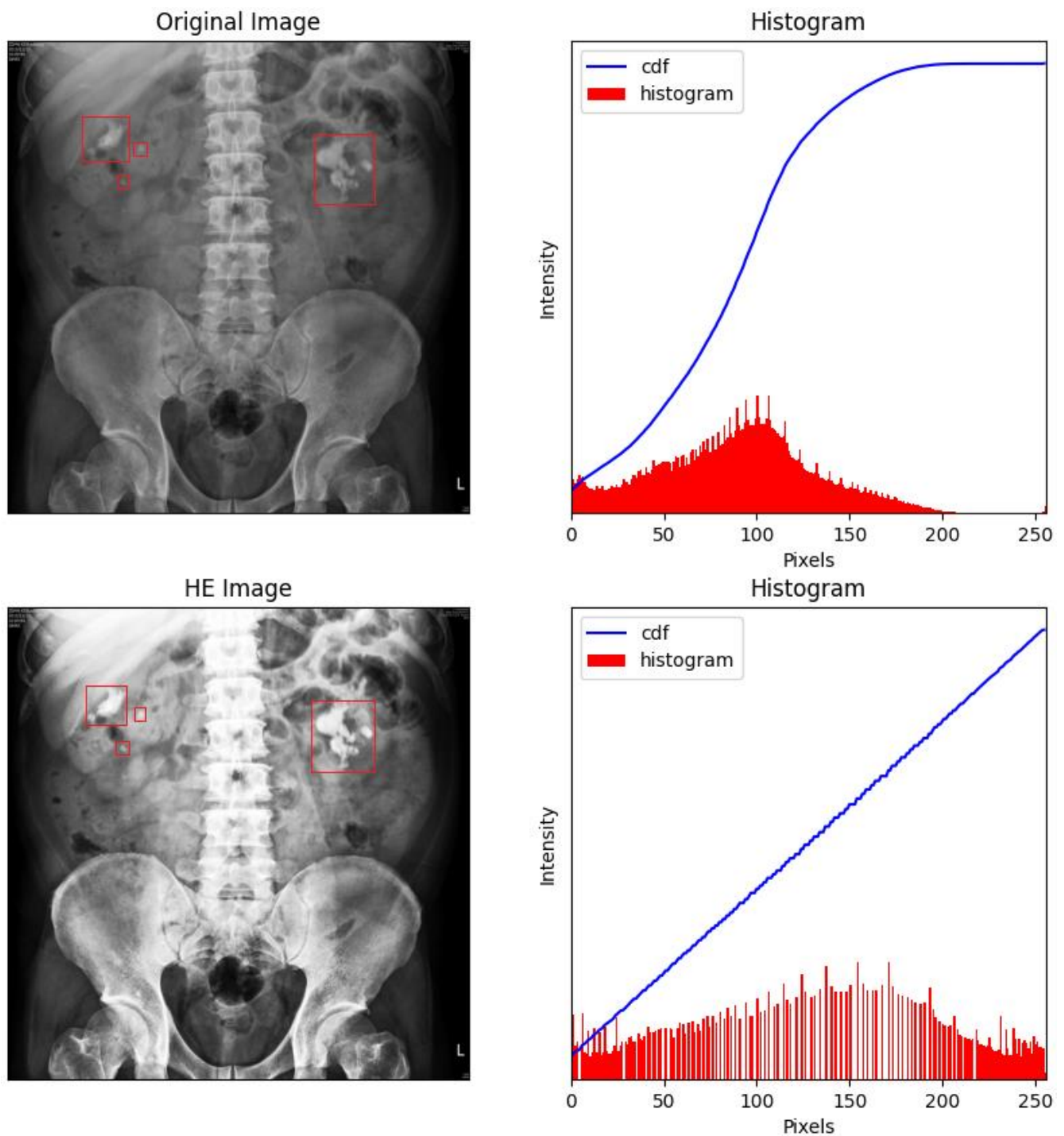
**Figure 4.** Confusion matrix and evaluation metrics.

### 3. Results

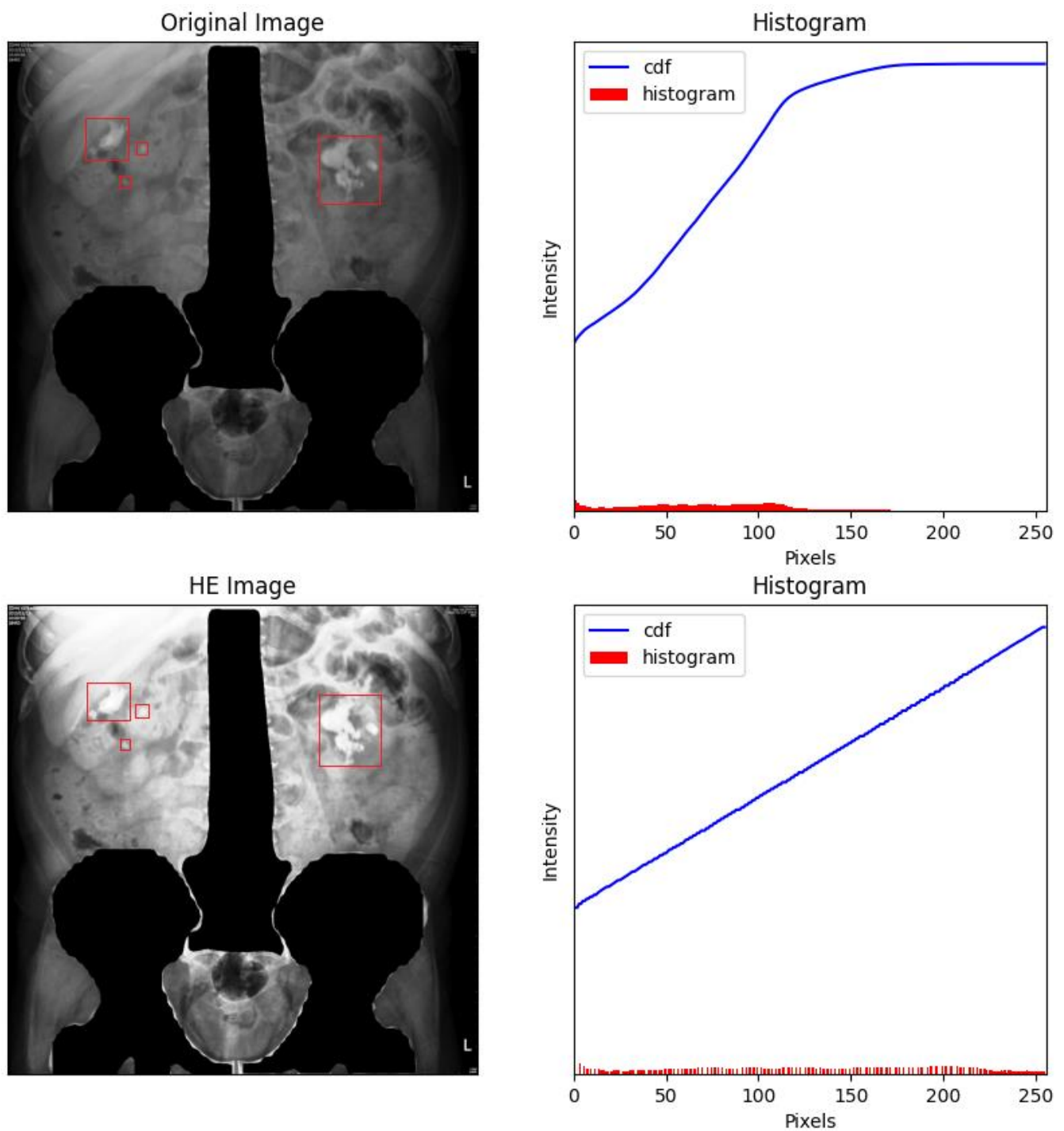
#### 3.1. Image Pre-Processing with Histogram

When histogram equalization (HE) is performed on KUB images, overexposure often occurs at the spine and pelvis, which tends to affect training negatively. Figure 5 shows an HE-processed image overexposed around dense tissues (bone), especially around the pelvis and spine, which may induce deviations during the feature extraction process. Therefore, all high-density regions in the KUB images must be masked. As shown in Figure 6, masking the spine and pelvis greatly reduced the high-intensity area of the images. Nonetheless, some overexposure still occurred at the rib cage, which is a common problem in HE. To prevent image overexposure from HE, CLAHE was performed on the KUB images. From Figure 7, it may be clearly observed that the CLAHE-processed image exhibits relatively little overexposure. Therefore, the CLAHE-processed KUB images were considered suitable for the observation of kidney stones.



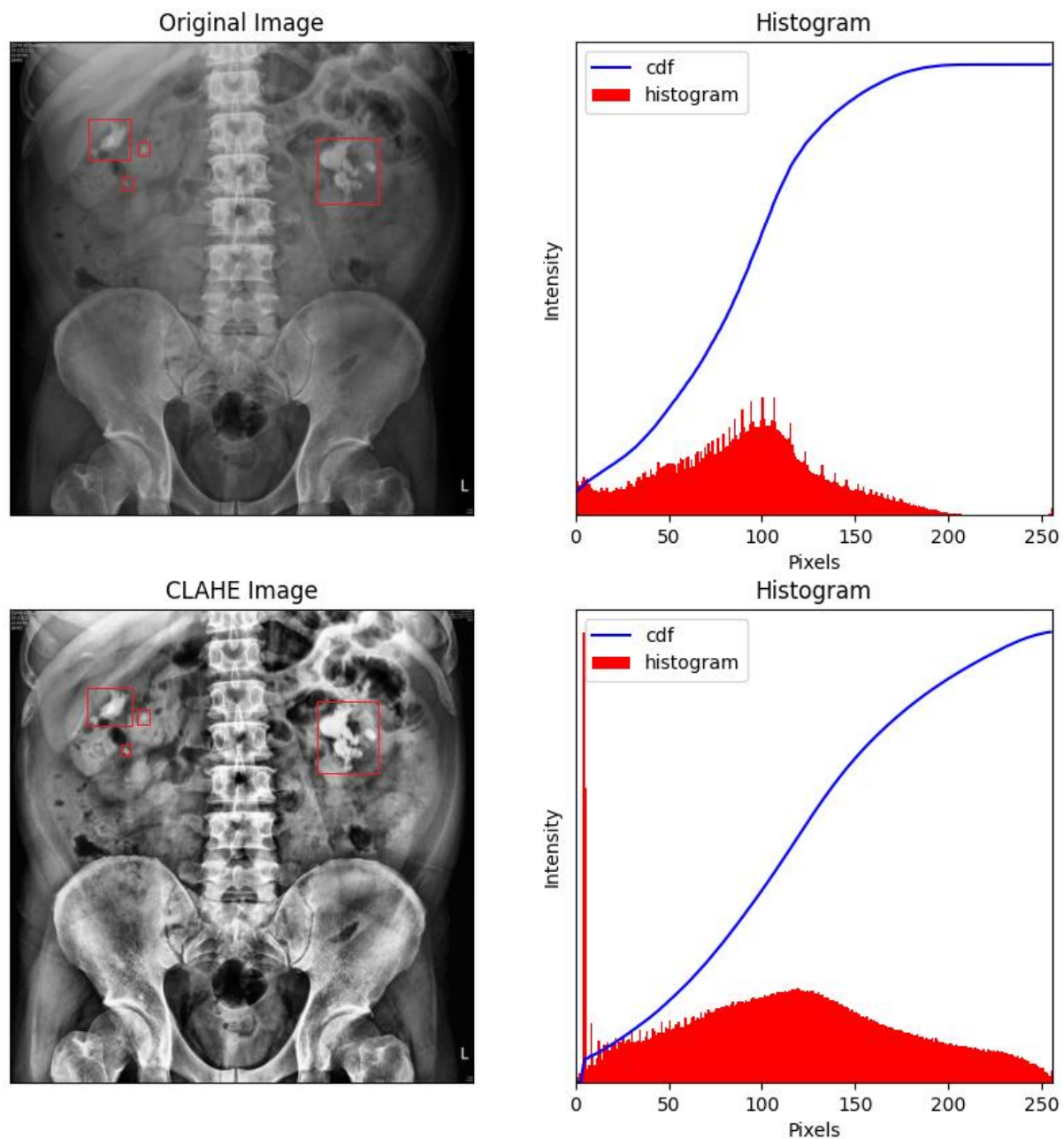


**Figure 5.** Many areas were overexposed in the HE-processed images, especially around dense tissue such as bone. The renal stones were labelled with red frames by the experts.



**Figure 6.** Masking the spine and pelvis greatly decreased the high-intensity areas of each image. The renal stones were labelled with red frames by the experts.



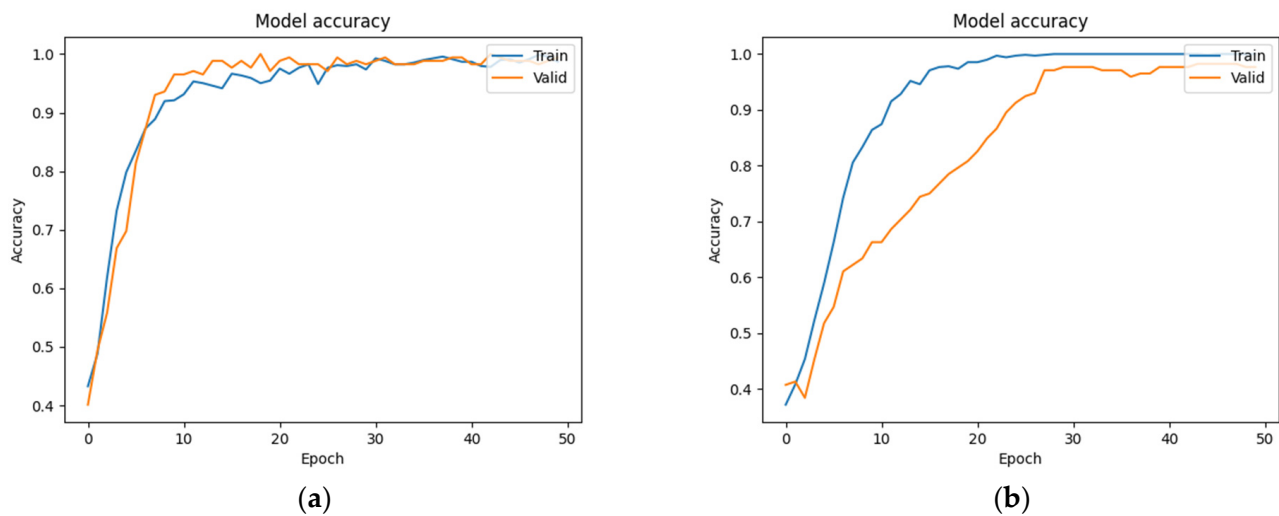


**Figure 7.** Contrast-limited adaptive histogram equalization (CLAHE) of a KUB image. It can clearly be observed that CLAHE greatly reduced overexposure around the rib cage, which makes identifying kidney stones relatively straightforward. The renal stones were labelled with red frames by the experts.

### 3.2. Effects of Data Augmentation on Training

We trained the ResNet model using both augmented and non-augmented datasets. Data augmentation was performed by rotating, horizontally and vertically translating, magnifying/shrinking, and shear-mapping the original images. In the augmented dataset (which contained the same number of images as the non-augmented dataset), these data augmentation procedures were randomly applied to every image after each iteration to ensure that the training data differed between iterations. The results were then compared in terms of accuracy and loss. Figure 8a,b show the results obtained with and without data augmentation, respectively. Although the accuracy increased much more rapidly when the

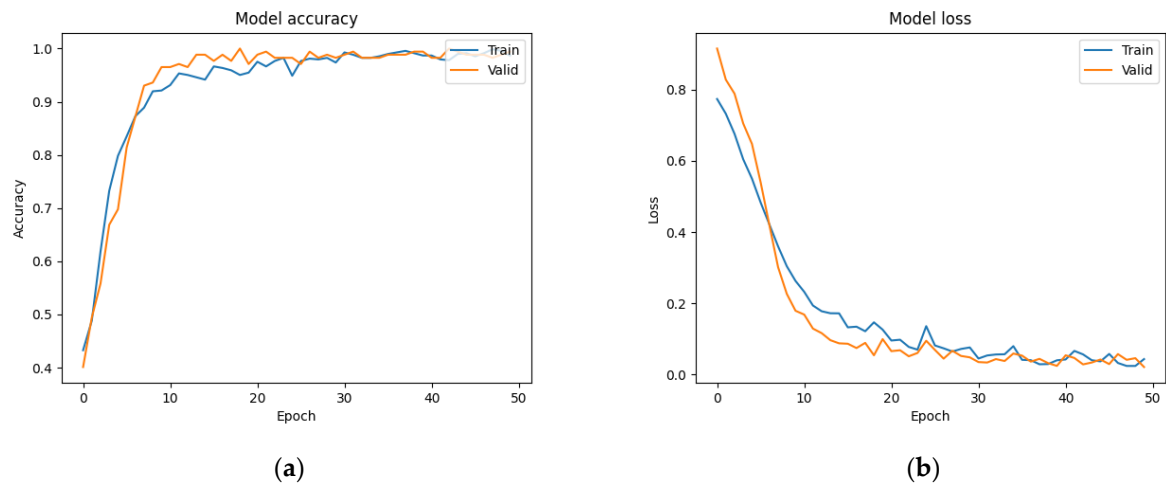
model was trained on the non-augmented dataset, it was unable to obtain a similar level of accuracy on the validation dataset in that case.



**Figure 8.** (a) Training accuracy using data augmentation. (b) Training accuracy without data augmentation.

### 3.3. Experimental Results

The model was trained for 50 epochs with an initial learning rate of  $10^{-5}$ . Because appropriate decreases to the learning rate are conducive for optimization, the learning rate was multiplied by 0.5 if the validation loss was not updated for five continuous epochs. The epoch-wise changes in accuracy and loss are shown in Figure 9a,b, respectively. It can be observed that the process of training from 0 epochs to 20 epochs converged rapidly. The accuracy and loss of the training set and the verification set were close, indicating that the model learned features in the initial stage well and classified them accurately. The subsequent loss from the 20th to the 50th epochs gradually converged to the optimal solution as the training ended. According to the confusion matrix shown in Table 1, the final accuracy of the model was 0.977, and its accuracy on the testing set was 0.982. Sensitivity is the ratio of patients with kidney stones who were correctly identified as positive cases, while precision is the ratio of correct diagnoses among positive cases. Therefore, a high sensitivity implies that false negatives are rare. Specificity is the ratio of patients without kidney stones who were correctly diagnosed as negative cases. Therefore, a model with a high specificity is unlikely to misdiagnose healthy subjects as positive cases. The F1-measure is the harmonic mean of recall (sensitivity) and precision, which summarizes the performance of a model. In kidney stone classification, the focus is on sensitivity, as the primary goal is to correctly identify patients who suffer from kidney stones. The sensitivity, specificity, precision, and F1-measure scores of our model were 0.953, 1, 1, and 0.976 on the validation set and 0.964, 1, 1, and 0.982 on the testing set, respectively (see Table 2). The ROC curves were also plotted to test the effectiveness of the model, and their AUCs were 0.995 and 1 on the validation and testing sets, respectively (Figure 10a,b). When  $AUC > 0.5$ , the classification performance of a classifier is better than random guessing, and the model has positive predictive value. The AUC value of our model was quite close to 1, which shows that the performance of our model was close to that of a theoretically perfect classifier, and it was effective in predicting positive samples correctly.



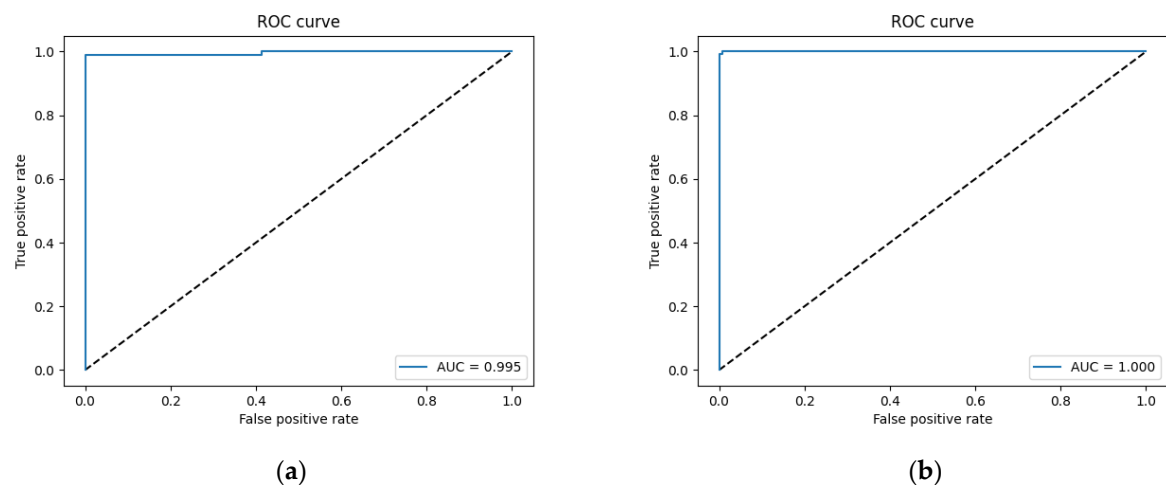
**Figure 9.** (a) Training accuracy and (b) training loss of ResNet model on our dataset.

**Table 1.** Confusion matrix of the model.

	TP	FP	TN	FN
Validation dataset	81	4	87	0
Testing dataset	132	5	137	0

**Table 2.** Performance of ResNet model on our dataset.

	Accuracy	Sensitivity	Specificity	Precision	F1-Measure	AUC
Validation dataset	0.977	0.953	1.000	1.000	0.976	0.995
Testing dataset	0.982	0.964	1.000	1.000	0.982	1.000



**Figure 10.** ROC curve of (a) validation set and (b) testing set.

### 3.4. Comparison of Accuracy with an Existing Method

The sensitivity, precision, and F1-measure of our method were 0.964, 1, and 0.982 on the testing set, respectively. Another CNN-based deep learning model [29] trained to detect kidney stones in pre-processed plain film X-ray images was also used for comparison, and the sensitivity, precision, and F1-measure of this model were 0.985, 0.762, and 0.862,

respectively, as shown in Table 3. Therefore, the performance of the proposed method was superior. This improvement may be attributed to the following factors. In addition to the differences in data collection, we utilized iterative data augmentation and various image pre-processing techniques. Data augmentation is commonly used in studies on medical imaging, especially to address overfitting with small datasets, and has achieved excellent results [46–48]. The use of CLAHE instead of HE for image pre-processing also helped reduce overexposure of the plain film X-ray images.

**Table 3.** Overall performance with CNN-based model [29].

	Sensitivity	Precision	F1-Measure
Proposed Model	0.964	1.000	0.982
CNN-based model [29]	0.985	0.767	0.862

#### 4. Discussion

In this study, we trained a CNN model to classify KUB images according to the presence of kidney stones. Although few studies have been conducted on the use of plain film X-ray images to detect kidney stones, the results are promising. According to a recent systematic review of recent AI advancements in urology by Dai et al. [49], only a single study used KUB images [29]. Other studies largely considered machine and deep learning models based on CT images, such as a work by Parakh et al. [50]. First, the advantages of plain film X-ray images include their low dose and cost, which enables them to be used in a wide range of medical institutions. Second, many deep learning models cannot accurately detect small objects or features, and kidney stones usually occupy an extremely small number of pixels in a KUB [51]. To address this problem, the images were cropped to enlarge the size of kidney stones and to train the model more easily. Third, the accuracy and generalizability of the model can be further improved by increasing the size of the training dataset. In the context of medical imaging, some plain film X-rays of kidney stones exhibit rarely-encountered patterns and features, which can make determining whether a kidney stone is present difficult. However, owing to their rarity, learning models cannot be trained on such images. By contrast, in most plain film X-ray images used to train the model, kidney stone(s) could be observed with the naked eye. If a large number of plain film X-ray images with difficult-to-observe kidney stones could be collected, the generalizability of the model could then be enhanced with further training to produce a highly reliable CAD tool. Although the kidney stones that our model was able to detect were obvious in the X-ray images, the model was nonetheless able to differentiate plain film X-ray images according to the presence or absence of such stones, which demonstrates that this approach can be extended to object detection and segmentation in the future. In deep learning studies on breast X-rays, over 4000 images have commonly been used to train deep learning models [52–55]. In this study, only 1130 images were used, and the small size of the dataset could have resulted in a poor training outcome. We therefore used data augmentation techniques to avoid severe overfitting and achieve adequate generalizability. In this study, we only used conventional data augmentation techniques as noted above rather than a generative adversarial network (GAN). Because GAN models have been successfully used to generate medical images [38,39,56,57], this approach remains as a potential direction for future research.

#### 5. Conclusions

In this study, we trained a ResNet model to classify KUB images based on the presence or absence of kidney stones. The proposed model presents excellent classification performance in terms of several metrics and can be used in the immediate diagnosis of kidney stones from plain film X-ray images. We draw the following conclusions from the results. (1) The retention of the spine and pelvis bones during image pre-processing exhibited an outsized impact on the accuracy of the model. (2) Overexposure from histogram

equalization reduced the accuracy and other evaluation metrics. This can be alleviated through masking and contrast-limited adaptive histogram equalization, which increase the training accuracy and improve the performance of the model. (3) Overfitting can be reduced for small datasets by augmenting the data used to train the model. This process also improves the generalizability of the model on unknown data, which explains why our model performed similarly on the validation and test sets. The proposed approach is expected to reduce the consumption of medical resources and limit the patients' radiation exposure, which is beneficial for both patients and physicians.

In the future, the proposed ResNet model could be combined with object detection or image segmentation strategies, such as SSD, Inception, or U-Net, to effectively detect very small kidney stones. In addition, we plan to consider topics beyond image classification. Once the classification model is more complete, we plan to study object detection and segmentation methods to locate and label any kidney stone appearing in KUB images, where each image may contain one or many objects of varying types. For object detection, we expect to adopt RetinaNet [58], which adds a single-shot multibox detector (SSD) to the frontend of ResNet and utilizes a focal loss function to improve image classification accuracy on unbalanced data, which is often the case for medical data. However, object detection methods only provide a rectangular bounding box enclosing a feature rather than the exact profile of an object, which can be crucial to diagnose a condition. Therefore, image segmentation is a quintessential part of an AI-driven CAD. To this end, we expect to use CaraNet as an image segmentation model [59]. In a  $1000 \times 1000$  px KUB image, a kidney stone may occupy a region smaller than  $20 \times 20$  px. Because CaraNet is specifically designed for the segmentation of small objects, we plan to study the feasibility of using CaraNet to improve the segmentation of small kidney stones in KUB images in future work.

**Author Contributions:** Conceptualization, Y.-Y.L., Z.-H.H. and K.-W.H.; methodology, Z.-H.H. and K.-W.H.; software, Z.-H.H. and K.-W.H.; validation, Y.-Y.L. and K.-W.H.; formal analysis, Y.-Y.L., Z.-H.H. and K.-W.H.; investigation, Y.-Y.L., Z.-H.H. and K.-W.H.; resources, Y.-Y.L.; writing—original draft preparation, Y.-Y.L., Z.-H.H. and K.-W.H.; writing—review and editing, Y.-Y.L. and K.-W.H.; visualization, Z.-H.H.; supervision, K.-W.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported in part by the Ministry of Science and Technology, Taiwan, R.O.C.; under grant MOST 110-2222-E-992-006-.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Romero, A.; Akpınar, H.; Assimios, D.G. Kidney stones: A global picture of prevalence, incidence, and associated risk factors. *Rev. Urol.* **2010**, *12*, e86–e96. [PubMed]
- Chewcharat, A.; Curhan, G. Trends in the prevalence of kidney stones in the United States from 2007 to 2016. *Urolithiasis* **2020**, *49*, 27–39. [CrossRef] [PubMed]
- Tundo, G.; Vollstedt, A.; Meeks, W.; Pais, V. Beyond prevalence: Annual cumulative incidence of kidney stones in the United States. *J. Urol.* **2021**, *205*, 1704–1709. [CrossRef] [PubMed]
- Bedel, C.; Uzun, A.; Korkut, M.; Kartal, M. Evaluation of modified stone score in patients presenting to the emergency department with flank pain. *Urol. Sci.* **2020**, *31*, 221. [CrossRef]
- Huang, H.-C.; Chou, Y.-W.; Chen, Y.; Liao, C.-H.; Chiang, B.-J. A lower urine white blood cell median can be a predictor of undiscovered urolithiasis in patients with acute urinary tract symptoms. *Urol. Sci.* **2020**, *31*, 115. [CrossRef]
- Niall, O.; Russell, J.; MacGregor, R.; Duncan, H.; Mullins, J. A comparison of noncontrast computerized tomography with excretory urography in the assessment of acute flank pain. *J. Urol.* **1999**, *161*, 534–537. [CrossRef]
- Wang, J.-H.; Shen, S.-H.; Huang, S.-S.; Chang, C.-Y. Prospective comparison of unenhanced spiral computed tomography and intravenous urography in the evaluation of acute renal colic. *J. Chin. Med. Assoc.* **2008**, *71*, 30–36. [CrossRef]



8. Fujii, K.; Aoyama, T.; Koyama, S.; Kawaura, C. Comparative evaluation of organ and effective doses for paediatric patients with those for adults in chest and abdominal CT examinations. *Br. J. Radiol.* **2007**, *80*, 657–667. [\[CrossRef\]](#)
9. Smith-Bindman, R.; Moghadassi, M.; Wilson, N. Radiation doses in consecutive CT examinations from five University of California Medical Centers. *Radiology* **2015**, *277*, 134. [\[CrossRef\]](#)
10. Metaxas, V.I.; Messaris, G.A.; Lekatou, A.N.; Petsas, T.G.; Panayiotakis, G.S. Patient doses in common diagnostic X-ray examinations. *Radiat. Prot. Dosim.* **2018**, *184*, 12–27. [\[CrossRef\]](#)
11. Brenner, D.J.; Hall, E.J. Computed tomography—An increasing source of radiation exposure. *N. Engl. J. Med.* **2007**, *357*, 2277–2284. [\[CrossRef\]](#)
12. Sagara, Y.; Hara, A.K.; Pavlicek, W.; Silva, A.C.; Paden, R.G.; Wu, Q. Abdominal CT: Comparison of low-dose CT with adaptive statistical iterative reconstruction and routine-dose CT with filtered back projection in 53 Patients. *Am. J. Roentgenol.* **2010**, *195*, 713–719. [\[CrossRef\]](#)
13. Ashour, A.S.; Dey, N.; Mohamed, W.S. *Abdominal Imaging in Clinical Applications: Computer Aided Diagnosis Approaches*; Medical Imaging in Clinical Applications; Springer: Cham, Switzerland, 2016; pp. 3–17.
14. Heidenreich, A.; Desgrandschamps, F.; Terrier, F. Modern approach of diagnosis and management of acute flank pain: Review of all imaging modalities. *Eur. Urol.* **2002**, *41*, 351–362. [\[CrossRef\]](#)
15. Panayides, A.S.; Amini, A.; Filipovic, N.D.; Sharma, A.; Tsaftaris, S.A.; Young, A.A.; Foran, D.J.; Do, N.V.; Golemati, S.; Kurc, T.; et al. AI in medical imaging informatics: Current challenges and future directions. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 1837–1857. [\[CrossRef\]](#)
16. Zhou, S.K.; Greenspan, H.; Davatzikos, C.; Duncan, J.S.; Ginneken, B.; Madabhushi, A.; Prince, J.L.; Rueckert, D.; Summers, R.M. A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. *Proc. IEEE* **2021**, *109*, 820–838. [\[CrossRef\]](#)
17. Lim, E.J.; Castellani, D.; So, W.Z.; Fong, K.Y.; Li, J.Q.; Tiong, H.Y.; Gadzhiev, N.; Heng, C.T.; Teoh, J.Y.-C.; Naik, N.; et al. Radiomics in urolithiasis: Systematic review of current applications, limitations, and future directions. *J. Clin. Med.* **2022**, *11*, 5151. [\[CrossRef\]](#)
18. Hameed, B.Z.; Shah, M.; Naik, N.; Khanuja, H.S.; Paul, R.; Somani, B.K. Application of artificial intelligence-based classifiers to predict the outcome measures and stone-free status following percutaneous nephrolithotomy for staghorn calculi: Cross-validation of data and estimation of accuracy. *J. Endourol.* **2021**, *35*, 1307–1313. [\[CrossRef\]](#)
19. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [\[CrossRef\]](#)
20. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.W.M.; van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [\[CrossRef\]](#)
21. Sarvamangala, D.R.; Kulkarni, R.V. Convolutional neural networks in medical image understanding: A survey. *Evol. Intell.* **2021**, *15*, 1–22. [\[CrossRef\]](#)
22. Fu, Y.; Lei, Y.; Wang, T.; Curran, W.J.; Liu, T.; Yang, X. Deep learning in medical image registration: A review. *Phys. Med. Biol.* **2020**, *65*, 20TR01. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Chan, H.-P.; Samala, R.K.; Hadjiiski, L.M.; Zhou, C. Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.* **2020**, *1213*, 3–21. [\[CrossRef\]](#)
24. Doi, K. Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Comput. Med. Imaging Graph.* **2007**, *31*, 198–211. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Chan, H.P.; Hadjiiski, L.M.; Samala, R.K. Computer-aided diagnosis in the era of deep learning. *Med. Phys.* **2020**, *47*, e218–e227. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Cha, K.H.; Hadjiiski, L.; Samala, R.K.; Chan, H.-P.; Caoili, E.M.; Cohan, R.H. Urinary bladder segmentation in CT urography using deep-learning convolutional neural network and level sets. *Med. Phys.* **2016**, *43*, 1882–1896. [\[CrossRef\]](#)
27. Långkvist, M.; Jendeborg, J.; Thunberg, P.; Loutfi, A.; Lidén, M. Computer aided detection of ureteral stones in thin slice computed tomography volumes using Convolutional Neural Networks. *Comput. Biol. Med.* **2018**, *97*, 153–160. [\[CrossRef\]](#)
28. Fitri, L.A.; Haryanto, F.; Arimura, H.; YunHao, C.; Ninomiya, K.; Nakano, R.; Haekal, M.; Warty, Y.; Fauzi, U. Automated classification of urinary stones based on microcomputed tomography images using convolutional neural network. *Phys. Med.* **2020**, *78*, 201–208. [\[CrossRef\]](#)
29. Kobayashi, M.; Ishioka, J.; Matsuoka, Y.; Fukuda, Y.; Kohno, Y.; Kawano, K.; Morimoto, S.; Muta, R.; Fujiwara, M.; Kawamura, N.; et al. Computer-aided diagnosis with a convolutional neural network algorithm for automated detection of urinary tract stones on plain X-ray. *BMC Urol.* **2021**, *21*, 1–10. [\[CrossRef\]](#)
30. He, K.; Gkioxari, G.; Dollr, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
31. Shen, W.; Xu, W.; Zhang, H.; Sun, Z.; Ma, J.; Ma, X.; Zhou, S.; Guo, S.; Wang, Y. Automatic segmentation of the femur and tibia bones from X-ray images based on pure dilated residual U-Net. *Inverse Probl. Imaging* **2021**, *15*, 1333. [\[CrossRef\]](#)
32. Zimmerman, J.; Pizer, S.; Staab, E.; Perry, J.; McCartney, W.; Brenton, B. An evaluation of the effectiveness of adaptive histogram equalization for contrast enhancement. *IEEE Trans. Med. Imaging* **1988**, *7*, 304–312. [\[CrossRef\]](#)
33. Pizer, S.M.; Amburn, E.P.; Austin, J.D.; Cromartie, R.; Geselowitz, A.; Greer, T.; ter Haar Romeny, B.; Zimmerman, J.B.; Zuiderveld, K. Adaptive histogram equalization and its variations. *Comput. Vis. Graph. Image Process.* **1987**, *39*, 355–368. [\[CrossRef\]](#)

34. Fawzi, A.; Samulowitz, H.; Turaga, D.; Frossard, P. Adaptive Data Augmentation for Image Classification. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 3688–3692.
35. Wang, J.; Perez, L. The effectiveness of data augmentation in image classification using deep learning. *Convolutional Neural Netw. Vis. Recognit.* **2017**, *11*, 1–8.
36. Shorten, C.; Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. *J. Big Data* **2019**, *6*, 60. [\[CrossRef\]](#)
37. Nanni, L.; Paci, M.; Brahnam, S.; Lumini, A. Comparison of different image data augmentation approaches. *J. Imaging* **2021**, *7*, 254. [\[CrossRef\]](#)
38. Frid-Adar, M.; Diamant, I.; Klang, E.; Amitai, M.; Goldberger, J.; Greenspan, H. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing* **2018**, *321*, 321–331. [\[CrossRef\]](#)
39. Ma, Y.; Liu, J.; Liu, Y.; Fu, H.; Hu, Y.; Cheng, J.; Qi, H.; Wu, Y.; Zhang, J.; Zhao, Y. Structure and illumination constrained GAN for medical image enhancement. *IEEE Trans. Med. Imaging* **2021**, *40*, 3955–3967. [\[CrossRef\]](#)
40. Wang, S.-Y.; Wang, O.; Zhang, R.; Owens, A.; Efros, A.A. CNN-generated images are surprisingly easy to spot for now. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 8695–8704.
41. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
42. Aswathy, R.H.; Suresh, P.; Sikkandar, M.Y.; Abdel-Khalek, S.; Alhumyani, H.; Saeed, R.A.; Mansour, R.F. Optimized Tuned Deep Learning Model for Chronic Kidney Disease Classification. *Comput. Mater. Contin.* **2022**, *70*, 2097–2111. [\[CrossRef\]](#)
43. Wright, L. Ranger—A Synergistic Optimizer. 2019. Available online: <https://github.com/lessw2020/Ranger-Deep-Learning-Optimizer> (accessed on 5 January 2022).
44. Liu, L.; Jiang, H.; He, P.; Chen, W.; Liu, X.; Gao, J.; Han, J. On the variance of the adaptive learning rate and beyond. *arXiv* **2019**, arXiv:1908.03265.
45. Zhang, M.; Lucas, J.; Ba, J.; Hinton, G.E. Lookahead optimizer: K steps forward, 1 step back. *arXiv* **2019**, arXiv:1907.08610.
46. Hussain, Z.; Gimenez, F.; Yi, D.; Rubin, D. Differential Data Augmentation Techniques for medical imaging classification tasks. *Annu. Symp. Proc.* **2018**, *2017*, 979–984.
47. Zhao, A.; Balakrishnan, G.; Durand, F.; Guttag, J.V.; Dalca, A.V. Data augmentation using learned transformations for one-shot medical image segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 8543–8553.
48. Chlap, P.; Min, H.; Vandenberg, N.; Dowling, J.; Holloway, L.; Haworth, A. A review of medical image data augmentation techniques for deep learning applications. *J. Med. Imaging Radiat. Oncol.* **2021**, *65*, 545–563. [\[CrossRef\]](#) [\[PubMed\]](#)
49. Dai, J.C.; Johnson, B.A. Artificial intelligence in endourology: Emerging technology for individualized care. *Curr. Opin. Urol.* **2022**, *32*, 379–392. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Parakh, A.; Lee, H.; Lee, J.H.; Eisner, B.H.; Sahani, D.V.; Do, S. Urinary stone detection on CT images using deep convolutional neural networks: Evaluation of model performance and generalization. *Radiol. Artif. Intell.* **2019**, *1*, e180066. [\[CrossRef\]](#) [\[PubMed\]](#)
51. Chen, C.; Liu, M.-Y.; Tuzel, O.; Xiao, J. R-CNN for small object detection. In Proceedings of the Asian Conference on Computer Vision, Taipei, Taiwan, 20–24 November 2016; pp. 214–230.
52. Islam, Z.; Islam, M.; Asraf, A. A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images. *Inform. Med. Unlocked* **2020**, *20*, 100412. [\[CrossRef\]](#)
53. Pathan, S.; Siddalingaswamy, P.; Ali, T. Automated detection of Covid-19 from chest X-ray scans using an optimized CNN architecture. *Appl. Soft Comput.* **2021**, *104*, 107238. [\[CrossRef\]](#)
54. Gazda, M.; Plavka, J.; Gazda, J.; Drotár, P. Self-Supervised deep convolutional neural Network for chest X-Ray Classification. *IEEE Access* **2021**, *9*, 151972–151982. [\[CrossRef\]](#)
55. Feng, Y.; Xu, X.; Wang, Y.; Lei, X.; Teo, S.K.; Sim, J.Z.T.; Ting, Y.; Zhen, L.; Zhou, J.T.; Liu, Y.; et al. Deep Supervised Domain Adaptation for Pneumonia Diagnosis From Chest X-Ray Images. *IEEE J. Biomed. Health Inform.* **2021**, *26*, 1080–1090. [\[CrossRef\]](#)
56. Al-Qerem, A.; Abu Salem, A.; Jebreen, I.; Nabot, A.; Samhan, A. Comparison between transfer learning and data augmentation on medical images classification. In Proceedings of the 22nd International Arab Conference on Information Technology (ACIT), Muscat, Oman, 21–23 December 2021; pp. 1–7. [\[CrossRef\]](#)
57. Waqas, N.; Safie, S.I.; Kadir, K.A.; Khan, S.; Khel, M.H.K. DEEPFAKE Image Synthesis for Data Augmentation. *IEEE Access* **2022**, *10*, 80847–80857. [\[CrossRef\]](#)
58. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
59. Lou, A.; Guan, S.; Ko, H.; Loew, M. CaraNet: Context axial reverse attention network for segmentation of small medical objects. *Image Process. SPIE* **2022**, *12032*, 81–92.