WILEY | Hindawi

Research Article

Exploring the Effect of Image Enhancement Techniques with Deep Neural Networks on Direct Urinary System (DUSX) Images for Automated Kidney Stone Detection

Ugur Kilic [ID],[1,2] **Isil Karabey Aksakalli** [ID],[1] **Gulsah Tumuklu Ozyer** [ID],[2] **Tugay Aksakalli** [ID],[3] **Baris Ozyer** [ID],[2] **and Senol Adanur** [ID][4]

[1]*Department of Computer Engineering, Faculty of Engineering, Erzurum Technical University, Erzurum, Türkiye*
[2]*Department of Computer Engineering, Faculty of Engineering, Ataturk University, Erzurum, Türkiye*
[3]*Department of Urology, Erzurum Regional Training and Research Hospital, Erzurum, Türkiye*
[4]*Department of Urology, Faculty of Medicine, Ataturk University, Erzurum, Türkiye*

Correspondence should be addressed to Ugur Kilic; ugur.kilic@erzurum.edu.tr

In diagnosing kidney stone disease, clinical specialists often apply medical imaging techniques such as CT and US. Among these imaging techniques, is frequently chosen as the primary examination method in emergency services due to its low cost, accessibility, and low radiation levels. However, interpreting the images by inexperienced specialists can be challenging due to the low image quality and the presence of noise. In this study, we propose a computer-aided diagnosis system based on deep neural networks to assist clinical specialists in detecting kidney stones using Direct Urinary System (DUSX) images. Firstly, in consultation with clinical specialists, we created a new dataset composed of 630 DUSX images and presented it publicly. We also defined preprocessing steps that incorporate image enhancement techniques such as GF, LoG, BF, HE, CLAHE, and CBC to enable deep neural networks to perceive the images more clearly. With these techniques, we considered the noise reduction in the DUSX images and enhanced the poor quality, especially in terms of contrast. For each preprocessing step, we created models to detect kidney stones using YOLOv4 and Mask R-CNN architectures, which are common CNN-based object detectors. We examined the effects of the preprocessing steps on these models. To the best of our knowledge, the combination of BF and CLAHE which is called CBC in this study, has not been applied before in the literature to enhance DUSX images. In addition, this study is the first in its field in which the YOLOv4 and Mask R-CNN architectures have been used for the detection of kidney stones. The experimental results demonstrated the most accurate method is the YOLOv4 model, which includes the CBC preprocessing step, as the result model. This model shows that the accuracy rate, precision, recall, and F1-score were found as 96.1%, 99.3% 96.5%, and 97.9% respectively in the test set. According to these performance metrics, we expect that the proposed model will help to reduce the unnecessary radiation exposure and associated medical costs that come with CT scans.

1. Introduction

The human excretory system consists of kidneys, ureters, and bladder. The system has a crucial role for human health. The kidneys in the excretory system filter toxic materials from the blood, particularly urea and the system ensures that they are eliminated from the body via the bladder [1]. Crystallized structures called "kidney stones" may occur in the kidneys when they perform their filtering function

within the body. Kidney stones are one of the most common ailments affecting the kidney and urinary system due to complications with the kidney's internal mechanism. Figure 1 shows an example of a kidney stone that occurred in the urinary tract and kidneys.

Small-sized kidney stones are thrown out by the urine without having any impact on the body. In contrast, as the diameter of the formed stones grows, they cause symptoms such as bloody urine, nausea, painful urination, and severe
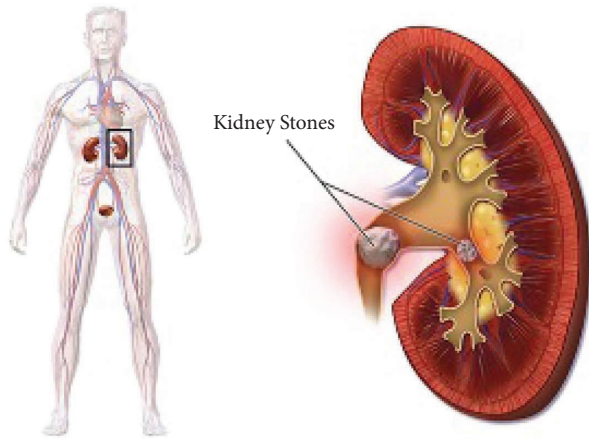
Figure 1: Kidney stones occurred in urinary tract [2].

lower abdomen or back pain. Patients suffer unbearable pain from kidney stones when they come out of the kidney and fall into the urinary canal. As the detection process of these kidney stones takes longer, the quality of life worsens, which leads to kidney function deteriorating and human life is endangered. Therefore, the diagnosis of kidney stones at an early stage is significant in the treatment process [3, 4]. Many patients who have kidney stone disease apply to hospitals with various clinical manifestations such as fever, severe pain in the lower back and sides, and blood in the urine [5]. In some cases, the disease is confused with clinics such as appendicitis, cholecystitis, ovarian torsion, and mesenteric ischemia [6]. Due to this multiplicity of differential diagnosis and the accompanying physician density in the emergency evaluation, physicians may misdiagnose kidney stones and overlook the diagnosis of kidney stones in patients presenting with milder symptoms. Therefore, physicians demand additional imaging such as computerized tomography (CT) containing intense radiation from patients to diagnose [7]. The analysis and interpretation of these medical images are manually and subjectively performed by physicians. Physicians may misinterpret medical images in a short time due to fatigue and poor quality of contrast form in medical images. According to statistics, human-induced misdiagnosis rate can reach 10–30% in medical image analysis [8]. To minimize the misdiagnosis problem, computer-aided diagnostic systems are proposed as practical approaches that can help physicians make a diagnosis. Hence, numerous neural network models such as artificial intelligence, machine learning, and deep learning models have been widely used to increase diagnostic accuracy in medical image analysis [9, 10]. Deep learning models particularly convolutional neural networks (CNNs) have recently become popular in medical image processing because high-level feature can be extracted from objects, after the training phase is completed [11].

In this study, a computer-aided diagnosis system is proposed to help physicians by automating kidney stone detection using CNN architectures through DUSX images. Because of its widespread use, the low amount of radiation compared to other imaging techniques, and the availability of imaging devices even in the simplest medical clinics, DUSX images were used in the study. Despite the widespread use of this imaging technique, as far as we investigated, a DUSX dataset has not been encountered in the literature. Our dataset was approved by the Ataturk University Faculty of Medicine Clinical Research Ethics Committee, and the dataset is publicly available for scientific studies. Hereby, we contributed to the literature by publishing a new DUSX dataset retrieved by Ataturk University Research Hospital in Erzurum, Turkey. Moreover, we investigated the effect of six image enhancement techniques (Gaussian Filtering (GF), Laplacian of Gaussian Filtering (LoG), Bilateral Filtering (BF), Histogram Equalization (HE), Contrast-Limited Adaptive Histogram Equalization (CLAHE), and Combination of BF and CLAHE (CBC)) to increase the accuracy rate of CNN models on DUSX images for automated kidney stone detection. After the preprocessing step, models were created using CNN-based object detectors YOLOv4 and Mask R-CNN architectures to automatically detect kidney stones on the images. Among the evaluated models, YOL0v4 model using CBC technique as a preprocessing step has the best performance with 96.1% accuracy on the test set. The developed computer-aided diagnosis system is ready for clinical application.

The main contributions of this study are summarized in the following:

(i) A new dataset was generated using unique DUSX images obtained from the hospital. We consider that this public dataset will pave the way for further research.

(ii) Based on our investigation, the CNN-based object detectors YOLOv4 and Mask R-CNN architectures are first applied to DUSX images for kidney stone detection.

(iii) The presence of noise in DUSX images and the poor quality, especially in contrast form, make it difficult to notice some details, and the lack of these details leads to reducing the accuracy of CNN models. To address this issue, the effect of various image enhancement techniques on kidney stone detection was investigated.

(iv) Among many image enhancement techniques, a hybrid filtering approach named CBC + YOLOv4 was proposed to detect kidney stones. This technique outperformed the other preprocessing methods.

The rest of the study has been detailed in the following. Section 2 discusses the related kidney stone detection methods in the literature. Section 3 describes the dataset and image labeling process, preprocessing steps, and CNN models used in this study. Section 4 shows the experimental results regarding different parameters and preprocessing techniques, and finally, Section 5 concludes the study.

2. Related Work

In recent years, many studies have been conducted to detect kidney stones on medical images in the literature. Medical imaging techniques such as ultrasound, DUSX, MRI, CT,

and color Doppler are used to diagnose kidney stone disease [12]. The main focus of the studies is based on the information on whether the stone is found in the image or not, and it has not been represented the boundary of the stones in visual results as we performed in our study. Based on our investigation, CNN-based object detectors have not been used to detect kidney stones in DUSX images. Moreover, an open-source dataset of DUSX images has not been found in the literature as far as we investigated. In this section, literature studies related to kidney stone detection were discussed according to the imaging type. The previous studies and their general features are summarized in Table 1.

Viswanath and Gunasundari [30] conducted kidney stone detection in ultrasound images. They used Gabor filtering and histogram equalization as preprocessing steps to eliminate the speckled noise in the ultrasound images and provide clarity in the image. Then, the preprocessed ultrasound image was segmented using level-set segmentation. After the segmentation process, wave sub-bands were used to detect energy levels in the areas removed from the kidney. Since the energy level in the region of the stone is different from the threshold value, the energy levels assisted to predict the location of the stone. Using energy level, they trained the network model created from Multilayer Perceptron and Back Propagation ANN. The authors stated that their system has 98.8% accuracy rate. In another study, Verma et al. [13] applied a median filter, Gaussian filter, and unsharp masking processes to clarify the stones in the ultrasonic images. Entropy-based segmentation was performed to find the stone area using morphological operations such as erosion and dilatation. They used some classification techniques, such as KNN and SVM. According to the experimental results, KNN has better accuracy rates than SVM. The authors stated that the KNN classification technique has better performance than SVM with 89% accuracy rate. Selvarani and Rajendran [15] proposed a metaheuristic SVM-based method to detect kidney stones on ultrasound images. They proposed the adaptive mean median filter approach to remove speckle noise from ultrasound images. Segmentation was performed using the K-means clustering algorithm. They extracted GLCM features for classification. According to the experimental results, metaheuristic SVM classified images with 98.8% accuracy. Eskandari et al. [16] proposed an expectation–maximization segmentation algorithm to detect kidney stones in ultrasonic images. Noise removal was performed on these images using the wavelet thresholding technique. Then, the authors used the expectation–maximization algorithm to segment kidney stones (renal calculi) in renal ultrasound images. They achieved 99.96% accuracy and 82.38% precision rate. However, the authors experienced that the computation time (58.02 s) was slower than the traditional algorithms. Khan et al. [14] proposed a speckle reduction approach to detect kidney stones on practical ultrasound (US) images using median filters and image segmentation techniques. A median filter was used to smooth images and reduce noise. Besides, a thresholding technique was used to make the segmentation process more robust. The approach has a 96.82% accuracy rate and 92.16% sensitivity on fifty test cases.

Längkvist et al. [9] developed a method using CNNs for the detection of kidney stones from CT. As a preprocessing step, smoothing was performed with a Gaussian filter. As a preprocessing step, a Gaussian filter was applied for smoothing. This method uses raw pixels on 3D volumes instead of feature extraction. The authors obtained 100% sensitivity and 2.68 false positives per patient in stone detection. Chak et al. [17] classified CT images as with stone and without stone using the support vector machine- (SVM-) based linear classifier. Before the classification phase, they used a neural network-based feature extraction method and applied a preprocessing step to filter the speckle noise on the CT images. They stated that their system obtained 95%–99% accuracy. Parakh et al. [18] used cascading CNN architecture for kidney stone detection in CT images. They created two CNNs structures identifying the urinary tract and detected the presence of stones. The authors obtained 95% accuracy, 94% sensitivity, and 96% specificity. Soni and Rai [12] also used CT images to detect kidney stones. Histogram equalization was used as a preprocessing step, and emboss was applied to calculate the differences in colors according to the directions. The support vector machine (SVM) classification method was applied to divide the vector space into two separate regions as stone-affected and healthy kidneys. The experimental results show that the test model has 98.71% accuracy rate. Cui et al. [19] developed a deep learning and thresholding-based model for kidney stone detection. In addition, they focused on scoring the detection on non-contrast CT images. They used 3D U-Net architecture as a deep learning method. The proposed model achieved a sensitivity of 95.9% in detecting stones larger than 2 mm in diameter. Yildirim et al. [20] performed kidney stone detection on CT images taken from 433 subjects. The created dataset consists of 1799 images by taking different cross-sectional CT images for each subject. A model was created by using xResNet-50 (cross-residual network) architecture, and the experimental results show that the accuracy rate was found by 96.82% on the test dataset. In another study that used CT images, Baygın et al. [22] aimed to classify patients who have kidney stones or not. They proposed a new classification network called ExDark19. They used the KNN algorithm as a classifier, and this classifier achieved 99.22% accuracy rate in the test data. Islam et al. [21] conducted the detection of three main kidney diseases (kidney stones, cysts, and tumors) on 12,446 CT images. They used vision transformers (EANet, CCT, and Swin transformers) and deep learning models (ResNet, VGG16, and Inception v3) to detect kidney diseases. The authors stated that the most accurate method was the Swin transformer, and the model in the test images had 99.30% accuracy rate in detecting the three types of kidney disease. Sabuncu et al. [24] used the Inception v3 model as a reference to detect kidney stones in CT images. In their study, a test accuracy of 98.52% was achieved in detecting kidney stones from CT images. Patro et al. [23] aimed to reduce redundancy in feature maps without convolution overlap by using a Kronecker-product-based convolution method instead of the traditional CNN-based deep learning network on the CT imaging model. The authors highlighted that the proposed method made the network

TABLE 1: The literature studies based on the kidney stone detection on medical images.

Author	Year	Image type	Methodology	Accuracy (%)
Verma et al. [13]	2017	US	KNN/SVM	89.0
Khan et al. [14]	2022	US	Speckle reduction approach	96.8
Selvarani and Rajendran [15]	2019	US	Metaheuristic SVM	98.8
Eskandari et al. [16]	2022	US	Expectation-maximization	99.9
Längkvist et al. [9]	2018	CT	CNN	Sensitivity: 100
Chak et al. [17]	2019	CT	ANN + SVM	95.0
Parakh et al. [18]	2019	CT	Cascading CNN	95.0
Soni and Rai [12]	2020	CT	SVM	98.7
Cui et al. [19]	2021	CT	3D U-Net and thresholding-based	90.3
Yildirim et al. [20]	2021	CT	xResNet-50	96.8
Islam et al. [21]	2022	CT	Transformers, ResNet, VGG16, Inception v3	99.3
Baygin et al. [22]	2022	CT	ExDark19	99.2
Patro et al. [23]	2023	CT	Deep kronecker neural network (DKN)	98.6
Sabuncu et al. [24]	2023	CT	Inception v3	98.5
Akshaya et al. [25]	2020	MRI	GLCM + BPN	—
EL-Sheikh et al. [26]	2022	MRI	Signal intensity	69.4
Parvathi et al. [27]	2023	MRI	3D-CU net model	91.2
Kobayashi et al. [28]	2021		Patchwise + ResNet	Sensitivity: 87.2
Preedanan et al. [29]	2023		U-Net pipeline	80
This study	**2023**		CBC + YOLOv4	**96.1**

efficient by extracting abstract and in-depth features from the input images. The method was validated using 10-fold cross-validation, and experimental studies show that the detection of kidney stones from CT images has an accuracy rate of 98.56%.

A method proposed by Akshaya et al. [25] used DWT (discrete wavelet transform) as a preprocessing step to detect kidney stones in the MRI images. Key features were extracted using GLCM (gray level co-occurrence matrix). The GLCM matrix determines the texture properties of an image by calculating how often pixel pairs occur with certain values in a specified spatial relationship. A dataset generated by 20 test data containing normal and abnormal kidney MRI images was classified using the backpropagation method. Kobayashi et al. [28] proposed a deep learning-based CAD (computer-aided diagnosis) system to detect kidney stones on plain images. They used 17-layer ResNet architecture for patchwise training. According to experimental results, their CAD system showed 87.2% sensitivity and 66.2% positive predictive value (PPV). Preedanan et al. [29] have proposed a two-stage pipeline for detecting kidney stones using the segmentation technique. Firstly, the location of urinary organs in images was identified using U-Net. Then, segmented images were increased using data augmentation methods and are passed through the second phase U-Net network to reduce class imbalance in the resulting map. Experimental results have shown that the U-Net network using the two-stage pipeline produces an accuracy of 80% in urinary stone classification.

3. Material and Method

3.1. Dataset and Image Labeling Process.
The generated dataset consists of 630 DUSX images obtained from patients who applied to Ataturk University's Urology Department, due to urinary system kidney stone disease. In the dataset,

558 images have one or more kidney stones in different regions with various sizes. The rest 72 images do not include a stone, and they are labeled as healthy kidneys. The dataset is split by ~ 80% as the training set and ~ 20% as the testing set. The training dataset is also divided into two parts as ~ 80% training and ~ 20% validation in order to increase the training success in itself. The hierarchy of DUSX images including 844 stones in total used for training, validation, and testing, is represented in Figure 2. The entire dataset can be accessible in this study's "Data Availability" section.

Labeling is a critical process for applying images supervised learning methods in artificial neural networks. In this study, the boundaries of the kidney stones for each image were acknowledged by a specialist doctor who works in the Urology Department. Then LabelImg for YOLOv4 and VGG Image Annotator [31] for Mask R-CNN were used to draw boundaries of the kidney stone in each image. In the YOLO tagging format, a ⟨image_name.txt⟩ file with the same name was created for each image file. Each .txt file contains object class, object coordinates, object height, and width information as annotations for the corresponding image file. At the end of the labeling process, all images in labeled format were saved, and txt tag files in YOLOv4 format containing the coordinate information of kidney stones (center of the stone, x, y coordinate, width, and height of the stone) in each labeled image were obtained. In Mask R-CNN tagging format, all image files were stored in a single .json file format. In the JSON file, the file names of the tagged images and the x, y coordinates of the points forming the polygons of the tagged objects are found in each image.

3.2. Preprocessing.
The use of CNN models has become popular in medical image processing since their high performance in the detection and classification of many diseases. These models are also widely used in image processing, and they can learn high-level properties about objects in
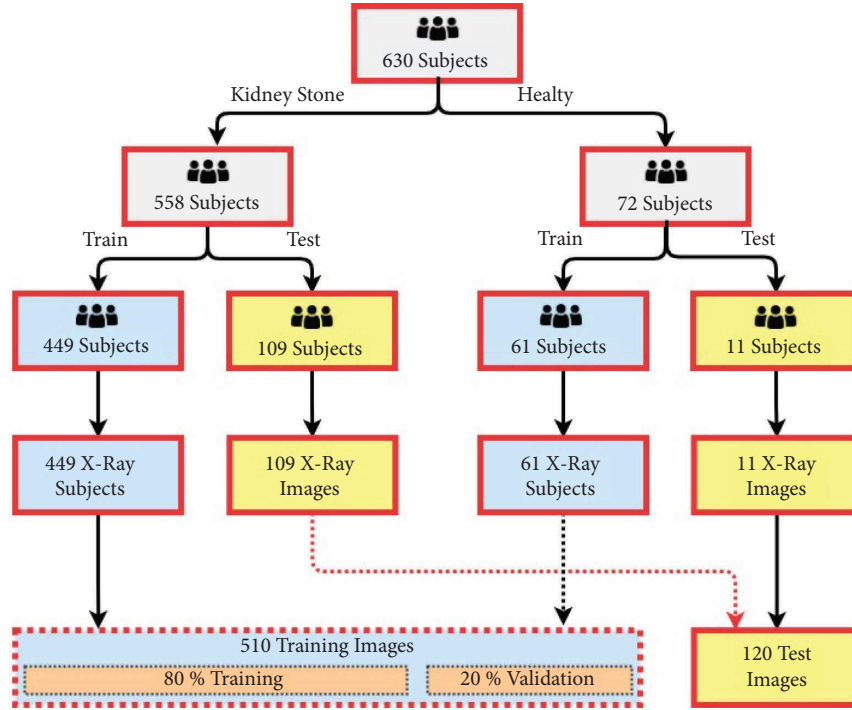
FIGURE 2: Data distributions used for training and testing phases in the study.

images by processing pixels in different layers. Thanks to these high-level features, the classification and detection operations can be performed successfully in the images. Before the classification and detection processes with CNN models, the accuracy of the models can be increased with different image enhancement techniques to be performed on the images.

In this study, object detection models were generated for automated kidney stone detection on DUSX images using CNN models. However, the presence of noise in DUSX images and the poor quality, especially in the form of contrast, make it difficult to notice some details and reduce the accuracy of the models. The images should be perceived more clearly by the models to increase the accuracy rates in the detection process. Therefore, various image enhancement techniques (GF, LoG, BF, HE, CLAHE, and CBC) were applied before generating CNN models and the effect of these techniques was investigated according to the experimental results. The coefficient values of the filters used in the preprocessing step are given in Table 2. Figure 3 demonstrates the original images and corresponding enhanced images applied in the preprocessing step. In Figure 3(a), kidney stones identified by the urologist were shown in red circles. The preprocessing phases performed in this study are described in the following subsections.

3.2.1. Gaussian Filtering (GF). The kernel of the filter is a discrete estimate of the normal distribution in the Gaussian filter. With this filter, an infinite transfer function can be filtered with a finite scanning window in the spatial domain. When a convolution operation is performed between an image with a Gaussian filter, the average of the pixels in the

image is considered and the difference in value between neighboring pixels is reduced. Hereby, noise is also reduced by smoothing the image. The Gaussian filter is generally used for operations such as noise removal, smoothing, and edge protection. The use of the Gaussian filter for two-dimensional images was represented in the equation (1) [32]. Equation (1) indicates the x, y values as the horizontal and vertical distances from the center, σ represents the standard deviation of the Gaussian distribution, and e represents the natural logarithm.

In this study, the window size of the Gaussian filter was chosen as 5×5 and σ value was 1, and DUSX images were smoothed at this rate. The large standard deviation value leads to larger peaks and this problem cause the image to be more blurred. Therefore, the value of σ was chosen as small as possible. Figure 3(b) shows the GF-applied images.

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\left(x^2 + y^2 / 2\sigma^2\right)}. \tag{1}$$

3.2.2. Laplacian of Gaussian Filtering (LoG). Laplacian is an operator representing linear quadratic derivative. The operator is used to define the edge transitions and contours in the images. Laplacian-based methods are sensitive to noises and when these methods are used on the images, the images have many unwanted edge points and noises. To handle this issue, the image is smoothed using Gaussian low-pass filtering in the LoG method [33, 34].

In this study, the LoG filter was applied to smooth the images, sharpen the edge contours of the kidney stones, and reduce the noise on DUSX images. LoG pixel values were calculated as shown in the following equation:

TABLE 2: Coefficients applied in filters.

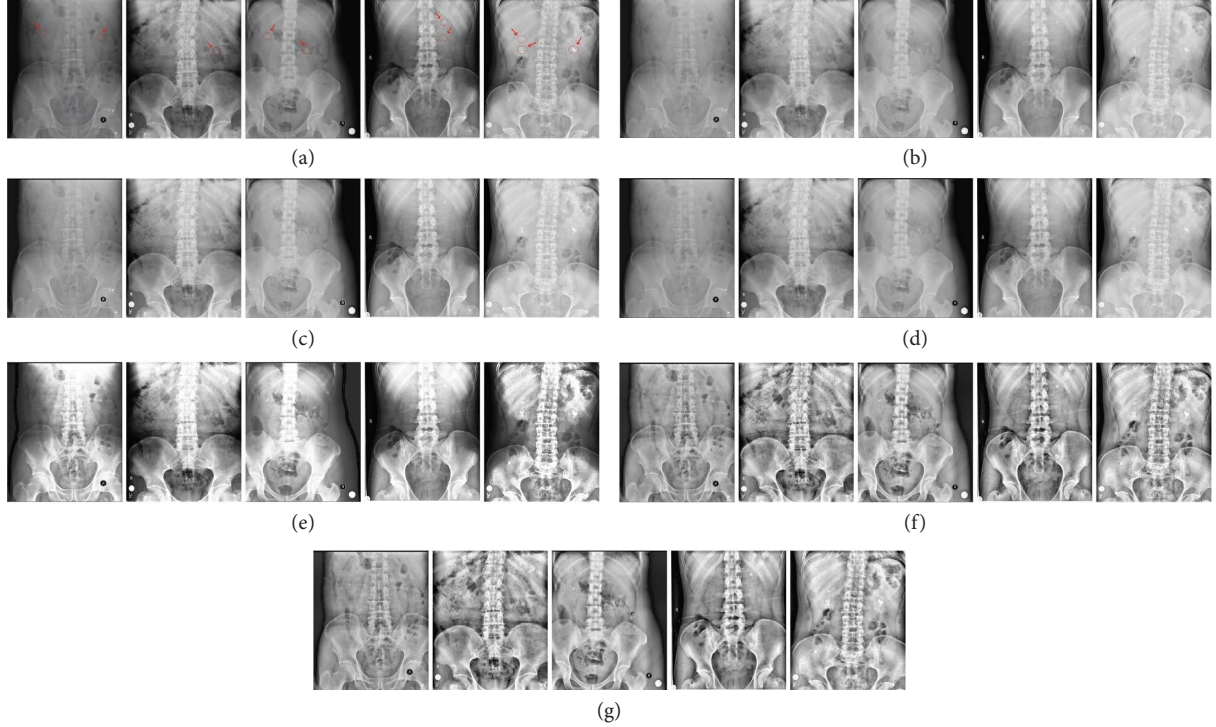| Method	Kernel size	σ	σ_s	σ_r	Clipping limit	Sub-block size
Gaussian filtering	5 × 5	1	—	—	—	—
Laplacian of Gaussian	5 × 5	1	—	—	—	—
Bilateral filtering	5 × 5	—	2	0.1	—	—
CLAHE	—	—	—	—	3	16
CBC	5 × 5	—	2	0.1	3	16

FIGURE 3: (a) Labeled original images. (b–g) Filtered images. (b) Gaussian filtering. (c) Laplacian of Gaussian filtering. (d) Bilateral filtering. (e) Histogram equalization. (f) CLAHE. (g) CBC.

$$\text{LoG}\ (x, y, z) = \frac{-1}{\pi\sigma^4}\left[1 - \frac{x^2 + y^2 + z^2}{2\sigma^2}\right]e^{-\left(x^2 + y^2 + z^2/2\sigma^2\right)}. \tag{2}$$

In Equation (2), (x, y) represents the pixels of the input image, σ is the standard deviation value, and LoG (x, y) represents the pixel values of the filtered image. The window size of the Gaussian filter was chosen as 5 × 5, σ value was selected as 1, and the input image was smoothed at this rate. A mask with edge pixel information of the image was obtained by applying the Laplacian operator on the image passed through the Gaussian filter. The output image was obtained by adding the original image and the mask containing the edge information. The original images, the mask, and the added version of the original image are shown in Figure 3(c).

3.2.3. Bilateral Filtering (BF). The bilateral filter is a basic antialiasing filter that aims to preserve edge information while smoothing images. Bilateral filters are frequently used when noise reduction is required by preserving the edge. BF includes a combination of two different Gaussian kernels, spectral and spatial kernels. Filtering is performed according to the spatial proximity of the central pixel to the neighboring pixels. In case of a high brightness difference between two pixels, it was aimed to preserve the sharp transition by adjusting the filter coefficient of the neighboring pixel according to the difference. In this way, the edge information of the image was preserved better than standard antialiasing filters during filtering. The BF process is expressed by the following equation:

$$BF[I]_p = \frac{1}{W_p}\sum_{q \in S}G_{\sigma_s}(\|p - q\|)G_{\sigma_r}\left(\left\|I_p - I_q\right\|\right)I_q. \tag{3}$$

In equation (3), I indicates the input image, p represents the currently filtered pixel position of the image, and q represents the neighboring pixels of the pixel that are in the S neighborhood. W_p normalization parameter is expressed by the following equation:

$$W_p = \sum_{q \in S}G_{\sigma_s}(\|p - q\|)G_{\sigma_r}\left(\left\|I_p - I_q\right\|\right). \tag{4}$$

The Gaussian kernel is expressed by the following equation:

$$G_\sigma(x) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right), \qquad (5)$$

where σ_s and σ_r indicate the standard deviation of the spatial smoothing function and the spectral effect function, respectively. The spatial kernel σ_s enhances the effect of nearby pixels, and the spectral kernel σ_r increases the effect of those with closer pixel values in the neighborhood. S is the area containing the neighborhoods centered on the p pixel in the image. The values of Gaussian kernels are the most effective factor for the filter performance. A high spectral kernel value causes the filter to execute like a typical Gaussian filter. As this value increases, the difference between the values of the pixels is not considered. Moreover, the increment of spatial parameter σ_s causes the smoothing larger features [35–37].

In the experiments conducted within the scope of this study, the best results were obtained using the following parameters on the bilateral filter: a window size of 5×5, σ_s 2 and σ_r 0.1. Since the σ values preserve the edge information, the kidney stones in the DUSX images can be distinguished better. Figure 3(d) shows bilateral filtering applied images.

3.2.4. Histogram Equalization (HE). Many contrast enhancement methods for medical images can be seen in the literature. Today, histogram equalization is one of the most preferred methods to improve the contrast of radiographic images. There are two types of classification for contrast enhancement methods in the literature. In the first type, contrast enhancement methods are divided into two classes according to frequency space and spatial space [38]. In the second type, contrast enhancement methods are classified as global and local methods. Global methods use the histogram of the entire image for contrast enhancement. As an alternative to the global method, local methods were developed to solve this problem. In native methods, the histogram of each subsection of the image is used instead of the entire image histogram in contrast enhancement [39]. In histogram equalization, the brightness distribution of the image is normalized to improve the global contrast of the image. Then, an output image with a uniform density distribution can be obtained. This operation is represented in the following equation [40]:

$$S(k) = \sum_{j=0}^{k} \frac{n_j}{n} * (L-1), k = 0, 1, 2, 3 \ldots, L-1. \qquad (6)$$

In equation (6), n_j is the number of pixels in the jth level, L is the desired gray level number (256 for 8 bits), and n is the total number of pixels. Figure 3(e) shows histogram equalization applied images.

3.2.5. Contrast-Limited Adaptive Histogram Equalization (CLAHE). The traditional histogram equalization method uses a global density distribution; for this reason, some important features can be suppressed by unimportant features such as background or noise [41, 42]. To solve this problem, the adaptive histogram equalization (AHE) method [43] was proposed in the literature.

In adaptive histogram equalization, the histogram of each sub-block of the image is used instead of the global histogram. Each pixel in each sub-block of the image is arranged in intensity proportional to the pixels in the surrounding region. In other words, the image is divided into sub-blocks in the form of a grid, and standard histogram equalization is applied to each sub-block. Then the sub-blocks are combined to obtain an enhanced image. Since the AHE method works in the local area, distortions called the blocking effect may occur in the border parts of the sub-blocks during the merging process. The blocking effect is a discontinuity problem. To solve this problem, the bilinear interpolation method is used for combining sub-blocks. Noise problems arise in local areas when contrast enhancement is performed with the AHE method. Especially noise increases in homogeneous regions. For this reason, the contrast-limited adaptive histogram equalization (CLAHE) method was developed to avoid noise by limiting the contrast enhancement [39, 43]. The CLAHE method helps to improve contrast in medical images without increasing the effect of noise [41].

In this study, we implemented CLAHE method to improve the contrast of DUSX images. In the CLAHE method, each image is divided into sub-blocks and the histogram of each sub-block is calculated. Each histogram is then cropped so that it does not exceed the clipping limit value (clipping limit = 3). Thus, the effectiveness of noises is prevented by limiting the contrast enhancement size. The number of clipped pixels is evenly distributed on the histogram. The histogram equalization method is applied for each histogram. In addition, the bilinear interpolation method is used to eliminate the blocking effect that may occur during the joining of sub-blocks. In the CLAHE method, local contrast enhancement is performed without increasing the amount of noise on the image. Figure 3(f) represents the CLAHE-applied images.

3.2.6. CBC. In this technique, noises are removed firstly by preserving the edge transitions by applying BF to the images. Then, CLAHE is applied to improve the contrast of the image, hereby, the local details can be identified easily. The results show that the use of the two combined methods is more beneficial than a uniform image enhancement. The general block diagram of the CBC method used in this study is given in Figure 4. In this method, the parameters were determined as the following: the window size of the bilateral filter is 5×5, σ_s 2 and σ_r 0.1. The sub-block size of the CLAHE method is 16×16, and the clipping limit is 3. Figure 3(g) shows CBC-applied images.

3.3. CNN Models. Determining the presence or absence of a disease in medical images is a wide-ranging application field, and it can be defined as a classification problem. In this field, deep learning methods, especially CNN algorithms present remarkable accuracy. In this study, YOLOv4 and
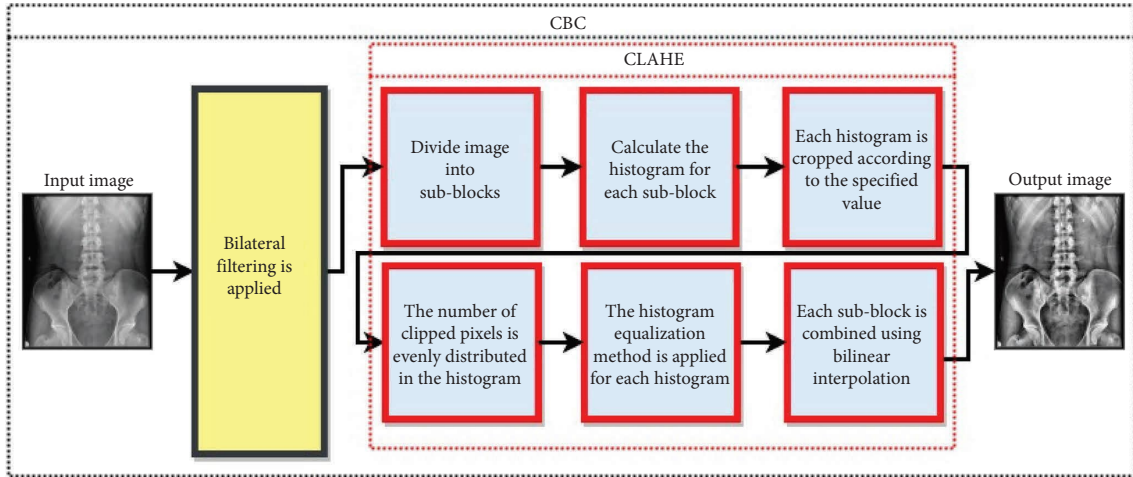
FIGURE 4: The general block diagram of the CBC method used in this study.

Mask R-CNN models were used for kidney stone detection in DUSX images. The training of the created CNN-based models was carried out on DUSX images. The general architectural structure used to compare different models and preprocessing methods applied in this study is shown in Figure 5. In the following subsections, we explained the architectural structures of YOLOv4 and Mask R-CNN models and the network configuration details for our dataset.

3.3.1. YOLOv4. In this study, the YOLOv4 model was used for the automatic detection of kidney stones in DUSX images. This section describes the YOLOv4 which is one of the most popular CNN-based object detectors. The network configurations for this model are explained in the following sections.

YOLO (You Only Look Once) [44, 45] is a CNN-based algorithm that can detect multiple objects in a single step with high accuracy and speed in real time with multibox structures. The YOLOv4 algorithm was proposed by Bochkovskiy et al. [46] in 2020 as the fourth version. The YOLOv4 version is an algorithm that can be trained quickly on a single graphics processing unit (GPU) and generates more accurate results than other versions. In the YOLOv4 model, the CSPDarknet53 [47] neural network is used as the feature descriptor. CSPDarknet53 performs splitting and merging operations on the feature map to provide more gradient flow from the Darknet-53 CNN. Darknet-53 is a convolutional network trained on ImageNet and it consists of 53 consecutive 1×1 and 3×3 convolutional layers followed by residual layers. Darknet-53 uses GPU efficiently due to the high number of floating-point operations it performs per second and makes the evaluation more efficient and faster than other feature extractors such as ResNet101 or ResNet152 [48, 49].

To extract features in the YOLOv4 architecture, SAM (spatial attention module), PAN (path aggregation network), and SPP (spatial pyramid pooling) structures [46] were implemented and their properties were extracted at three different scales to recognize objects of various sizes. When the input image is given to the network, the third scale divides the image into 52×52 cells, enabling the detection of small-sized objects. The second scale divides the image into 26×26 cells, allowing common-sized objects to be detected. The first scale allows the detection of large objects by dividing them into 13×13 cells. Using these sizes, the output size of each scale was calculated as $N \times N \times [3 \times (4 + 1 + C)]$. In this equation, the expression 3 represents the number of bounding boxes calculated for each cell, 4 denotes the offset values (t_x, t_y, t_w, t_h) of each bounding box, 1 denotes the objectivity score, and C the number of its class. Finally, the output of the network shows the boundary box's coordinates belong to the estimated object, the objectivity score, and the class information of the object [45, 48]

(1) Bounding Box Prediction. Anchor boxes are used to estimate the boundary boxes as shown in Figure 6. The best anchor boxes are calculated by applying the K-means clustering algorithm. When the K-means clustering algorithm is applied, the Intersection over Union (IoU) score is used instead of the Euclidean distance. If several anchors overlap, any anchor can be selected with the IoU value. The sizes of anchor boxes obtained by the K-means clustering algorithm are appropriately assigned to the scales. The network predicts four values for each bounding box (t_x, t_y, t_w, t_h). Using the sigmoid function, the center coordinate values (t_x, t_y) are reduced to the range 0-1. Using the equations in Figure 6, the center point of the anchor box obtained using K-means is calculated by its distance from the upper left corner of the grid cell $\sigma(t_x)$ and $\sigma(t_y)$. By adding the distances (c_x, c_y) to the upper left corner of the image to these values, the center point coordinates of the boundary box are found (b_x, b_y). Finally, the width and height of the boundary box (b_w, b_h) are calculated using the anchor box dimensions (p_w, p_h) using K-means. With the operation, $p_w e^{t_w}, p_h e^{t_h}$ the signs of the values are converted to positive in case of encountering negative t_x and t_y values. The YOLO network estimates an objectivity score for each anchor using logistic regression. The objectivity score
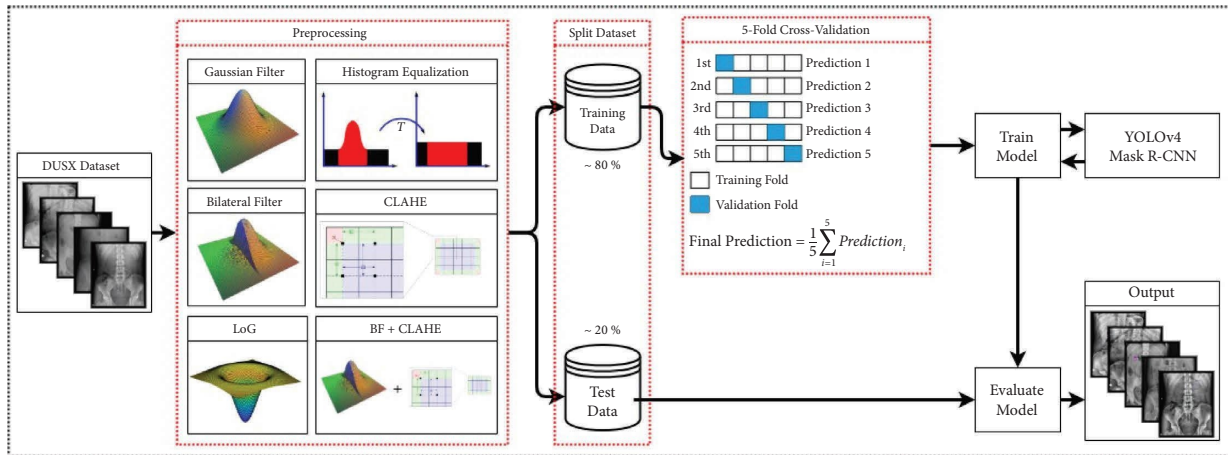
FIGURE 5: General architectural structure used in comparison to different preprocessing methods and CNN models for automated kidney stone detection (AKSD).
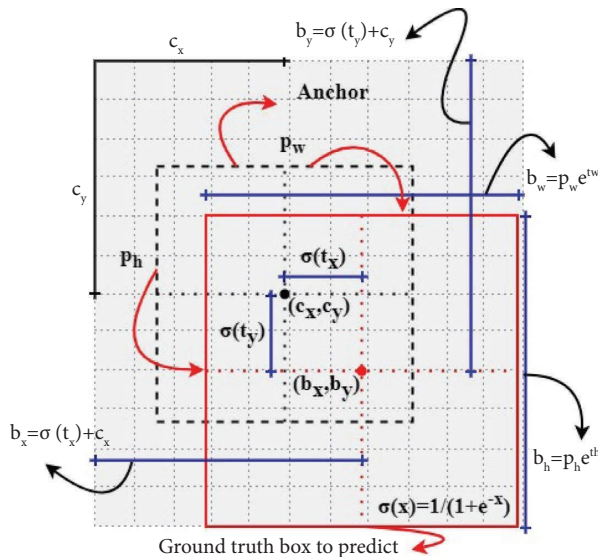

FIGURE 6: Estimation of bounding box using YOLO model [45].

estimate indicates the probability that the anchor contains an object. It also estimates the class probability using an independent logistic classifier. Class probability estimation performs object classification.

3.3.2. Mask R-CNN. In this study, the Mask R-CNN method evolved by region-based convolutional object detection algorithms was used to detect kidney stones in DUSX images. This section briefly introduces Mask R-CNN architecture. The network configuration details according to our dataset were explained in the following sections. Mask R-CNN is a neural network model which is based on object segmentation and developed by the Facebook Artificial Intelligence Research (FAIR) team as an extension of the Faster R-CNN algorithm. Instead of semantic segmentation identifying each pixel in the image, Mask R-CNN uses instance segmentation by segmenting only the pixels where the target objects are located and placing a mask on them. The

output of the network consists of the bounding box which is the output of the Faster R-CNN algorithm, and a mask defining the object outlines at the pixel level in addition to the object class. Region-based convolutional object detection algorithms generally consist of two parts: (1) identifying regions that could potentially be objects in the processed image and (2) delimiting the detected object with a bounding box [50, 51].

(1) Region Proposal Network (RPN). The algorithm gives the input image to the backbone (ResNet101), which is a standard convolutional boundary network that acts as a feature extractor. Convolutional feature mapping is generated by passing the image through the feature pyramid network in the backbone. The $n * n$ size window is shifted over the feature map; then this window is matched with a lower-dimensional feature vector. The RPN proposes region anchor boxes, which can be a set of objects, with different aspect ratios at each floating window location. Each proposed anchor box is associated with the box's objectivity score and the four coordinates of the bounding box. If the anchors highly overlap each other, the nonmax suppression (NMS) process is performed by choosing the one with the highest intersection-to-union (IoU) ratio value. Then intersection regions (Region of Interest, RoI) are obtained by NMS processing. [52].

(2) Region of Interest (RoI). Classifiers are capable of processing fixed-size input images better than variable-size input images. However, the RoI regions have different sizes due to the different aspect ratio bounding boxes in the RPN. Therefore, a process of resizing region proposals named RoI pooling is required to make regions fixed-sized. The RoI Align method was developed for the pooling process in the Mask R-CNN method [50].

In the RoI Align method, the region suggested by the RPN is divided into $n \times n$ grids. Because each grid cell is expected to contain the same number of pixels, fractional pixel states may occur. Then, each cell of the grid is sampled by dividing it into four subcells. Bilinear interpolation is

performed to represent subcells with a single value. In the last step, maximum pooling is performed on the bilinear interpolation values to obtain the $n \times n$ sized output. Fixed-size RoIs are sent to the fully connected layers to generate classification and bounding box information, and then they are sent to the mask branch to generate mask information.

(3) RoI Classifier and Bounding Box Regressor. In the RPN network, binary classification is performed as foreground/background for RoI's. Unlike RPN, the RoI classifier has a deeper network and it performs multiple classifications for each RoI. A bounding box is created for RoI in RPN. In the bounding box regressor, the bounding box is optimized to fully cover the object.

(4) Mask Branch. The mask branch is a CNN that uses the RoI information obtained after the RoI align procedure and it creates masks for them. The created masks have a lower resolution than 28×28 pixels. The small-sized mask ensures that the processing density is low and the mask branch remains lightweight. During the training phase, the basic truth masks are scaled to 28×28 size to calculate the value of the loss function. During the estimation, the masks are scaled to the dimensions of the RoI bounding box and final masks of the detected objects are created including one for each object [53].

(5) Loss Function. The loss/error function of the Mask R-CNN method is formalized in equation (7). The equation shows that the loss function consists of three sub-components. These subcomponents are the loss function for each classification result (L_{cls}), the loss function for the regression process used to determine the bounding box (L_{reg}), and the loss function for the segmentation mask (L_{mask}), respectively. The minimization of this function is performed iteratively using the gradient descent algorithm [50]:

$$L_{loss} = L_{cls} + L_{reg} + L_{mask}. \qquad (7)$$

3.3.3. Network Configurations and Model Training Phases. In YOLOv4 and Mask R-CNN constructs, network configuration files need to be adjusted for model training. These files contain many parameters such as the architectural structure, number of layers, activation functions, learning rate, and input image for network training and testing phases. In this study, all training processes were carried out on the Intel(R) Core(TM) i5-7400 CPU 3.00 GHz processor and Nvidia Geforce RTX 2080 8 GB GPU computer. To perform the training operations on the graphics card, the 10th version of the CUDA library was used to offer parallel computing on the graphics card created by Nvidia. Besides, OpenCV Library, which can be used as open source for operations on images, Keras, and TensorFlow libraries were implemented to train the Mask R-CNN model. Predictions were performed with the images reserved for verification, and the models were saved in the backup folder specified in the configuration file. k-Folds (5-fold) cross-validation was

performed in the training phase to ensure the randomness of the generated models and avoid the overfitting problem. As a result of cross-validation techniques, the average of five models was considered as a result model and the evaluation process was performed on this model. Different training processes were carried out for each preprocessing step applied to the images.

(1) Configuration and Model Training for YOLOv4. The parameters required for YOLOv4 model training were located in the "YOLOv4-custom.cfg" configuration file in the root directory of the Darknet backbone. The parameters shown in Table 3 were arranged and the configuration file was prepared for the training. After parameter and file configurations, initial weight values were identified randomly during the training phase. For this reason, the training period may be time consuming. To shorten the training time and transfer learning, Darknet-53 convolution weights, which were previously trained in the Imagenet, were identified as initial weights in the training of the "darknet53conv.74" YOLOv4 network. Each training process was performed in 6000 iterations and takes ~ 15 hours. Evaluation of the output models obtained as a result of the training phases is detailed in Section 4.

(2) Configuration and Model Training for Mask R-CNN. Matterport [54] Mask R-CNN which is one of the popular frameworks was chosen to train the Mask R-CNN model. The parameters required for model training were configured in the "conFigure py" file which is located in the root directory. Network configuration and training parameters for the Mask R-CNN model were identified in Table 4. We used the "mask_rcnn_coco.h5" file, which was previously trained with the Microsoft COCO [55] dataset as the initial weight values for the model training. Mask R-CNN takes a parameter named "layers" during the training phase. This parameter has two options named "heads" and "all." The "heads" option changes only the weights of the last layers of the model where transfer learning is performed, and the available weights of the Backbone and RPN (region proposal network) networks are not changed. When model training is performed by selecting the "all" option, all layers including Backbone and RPN are trained according to the new dataset. It was observed that a more successful model is identified as a result of the training by choosing "all" option. Each training process was performed at 400 epochs and takes ~ 10 hours. Evaluation of the output models obtained as a result of the training phases is detailed in Section 4.

4. Experimental Results

In this study, the evaluation process was carried out with the IoU (Intersection over Union) metric. The IoU is an evaluation metric that measures the similarity between the ground truth bounding box (labels denoted by the bounding boxes in the test dataset) and the predicted bounding box to evaluate the robustness. So, it is defined as the intersection of the junction of the detection bounding box and the ground truth bounding box. The IoU score ranges from 0 to 1, the

TABLE 3: YOLOv4 model training parameters.

Parameter	Definition	Value
Batch	Number of images to be processed at once	64
Subdivisions	Number of steps to process the batch value image	64
Width	The width of the resized input image	416
Height	The height of the resized input image	416
Max_batches	Maximum number of epoch	6000
Steps	Update parameter of learning rate (80% of Max_batches/90% of Max_batches)	4800/5400
Class	The number of class	1
Filters	Filters = (classes + coords + 1) ∗ 3	18

TABLE 4: Mask R-CNN model training parameters.

Parameter	Definition	Value
Images_Per_GPU	Number of images to train with on each GPU	1
Num_Classes	Number of classes (including background)	2
Steps_Per_Epoch	Number of training steps per epoch	100
Backbone	Backbone network architecture	ResNet101
NMS_Threshold	Nonmaximum suppression threshold for detection	0.5
MIN_Confidence	Minimum probability value to accept a detected instance	0.5
Epoch	Number of steps in training	400

closer the two boxes are, the higher the IoU score. A threshold value with a real value of 0.5 is commonly accepted to convert each object detection into classifications. If the kidney stone is detected according to IoU ≥ 0.5 threshold value, the object is classified as true positive (TP). When a label is present in the image and the model fails to detect the kidney stone, the object is classified as false negative (FN). If the image has no labeling but the image has a detection with IoU ≥ 0.5, it means that this is a false detection and it should be classified as false positive (FP). Using these parameters (TP, FP, TN, and FN), accuracy rate, precision, recall (sensitivity), F1-score, and specificity were calculated to compare the performances of the models. The equations of these evaluation metrics are represented in the following equations, respectively:

$$\text{Accuracy rate} = \frac{TP + TN}{TP + TN + FP + FN}, \quad (8)$$

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (9)$$

$$\text{Recall (sensitivity)} = \frac{TP}{TP + FN}, \quad (10)$$

$$F1 - \text{score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (11)$$

$$\text{Specificity} = \frac{TN}{TN + FP}. \quad (12)$$

The performance of the obtained models after the training phase was tested on 120 test images containing 142 stones. Initially, the models were tested on images without any preprocessing steps, and six preprocessing steps were applied to the images. Test images of YOLOv4 and Mask R-CNN models are given in Figures 7 and 8, respectively.

When the test images are examined, it is seen that the YOLOv4 model using the CBC preprocessing step was able to detect all stones in the figure.

Confusion matrix is a table of predictions and actual values used to evaluate the performance of classification models in machine learning. Confusion matrices with four combinations (TP, FP, FN, and TN) are created using estimated and real values. Figures 9 and 10 represent the confusion matrices for YOLOv4 and Mask R-CNN models, respectively.

When the confusion matrices of the models are examined, it is seen that the YOLOv4 model without applying the preprocessing step can detect 118 of 142 stones, and the Mask R-CNN model can detect 115 of them. The performance of the models increases in kidney stone detection by applying preprocessing steps. In particular, the model obtained by using the CBC method and the YOLOv4 model was able to detect 137 out of 142 stones. The accuracy rate, precision, recall, F1-score, and specificity values of the models were evaluated using confusion matrices to compare the performances of the models, and the results are shown in Table 5. When the performance of the models is evaluated according to the calculated metrics, it is seen that the application of BF, GF, LoG, CLAHE, and CBC preprocessing steps increased the accuracy of CNN models. On the other hand, HE did not significantly increase the accuracy of the models.

Kidney stone disease classification was performed in DUSX images using transfer learning method with EfficientNet [56], Densenet [57], ResNet101 [58], and MobileNet [59] deep learning architectures. The best 1000 features for each transfer learning method were selected with the relief algorithm, and the obtained features were classified by SVM using 5-fold cross-validation. Accuracy rate, precision, recall, and F1-score values of the models are given in Table 6. When Table 6 is examined, it can be seen that the
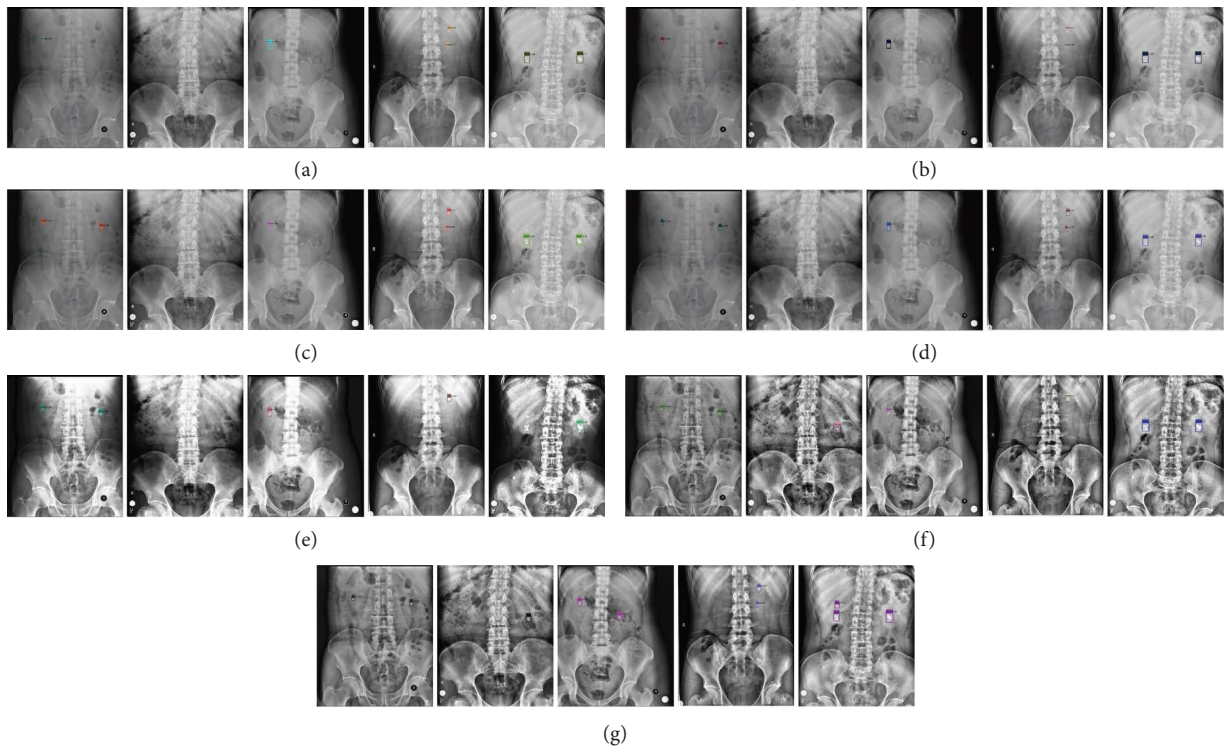
(a)

(b)

(c)

(d)

(e)

(f)

(g)

FIGURE 7: The result images of YOLOv4 models. (a) No preprocessing. (b) Gaussian filtering. (c) Laplacian of Gaussian filtering. (d) Bilateral filtering. (e) Histogram equalization. (f) CLAHE. (g) CBC.
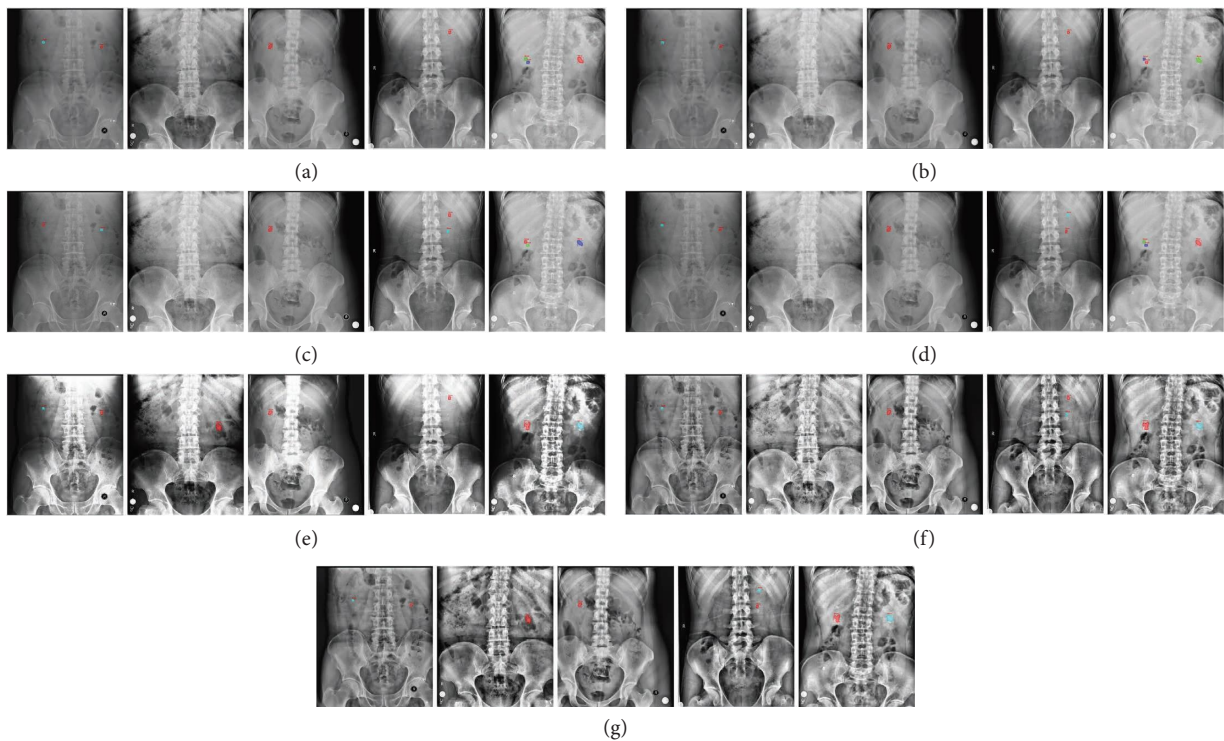

(a)

(b)

(c)

(d)

(e)

(f)

(g)

FIGURE 8: The result images of mask R-CNN models. (a) No preprocessing. (b) Gaussian filtering. (c) Laplacian of Gaussian filtering. (d) Bilateral filtering. (e) Histogram equalization. (f) CLAHE. (g) CBC.
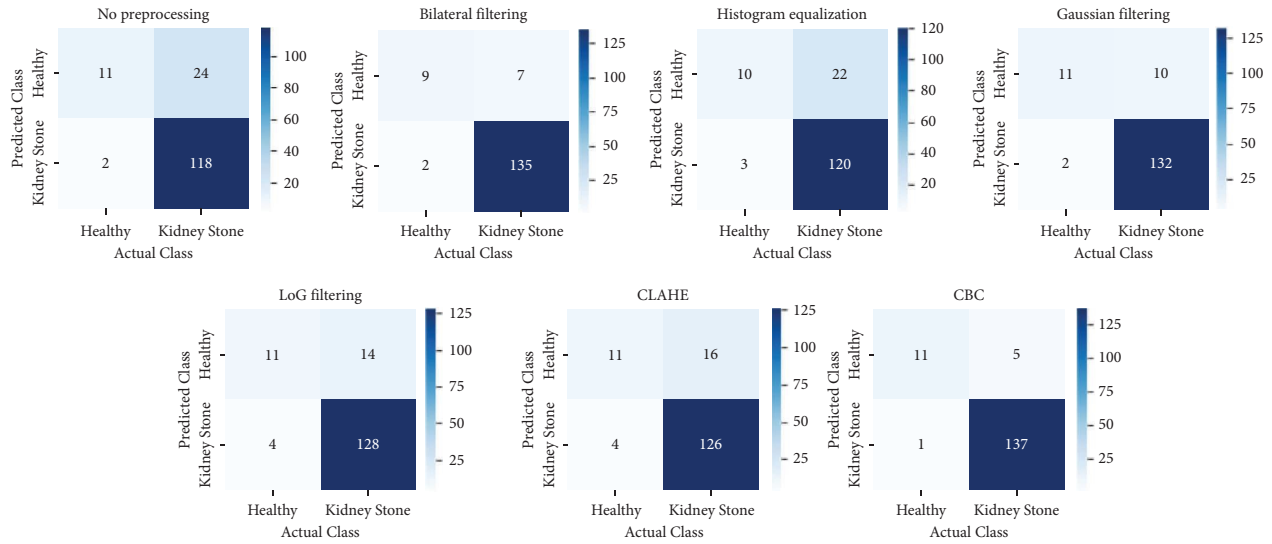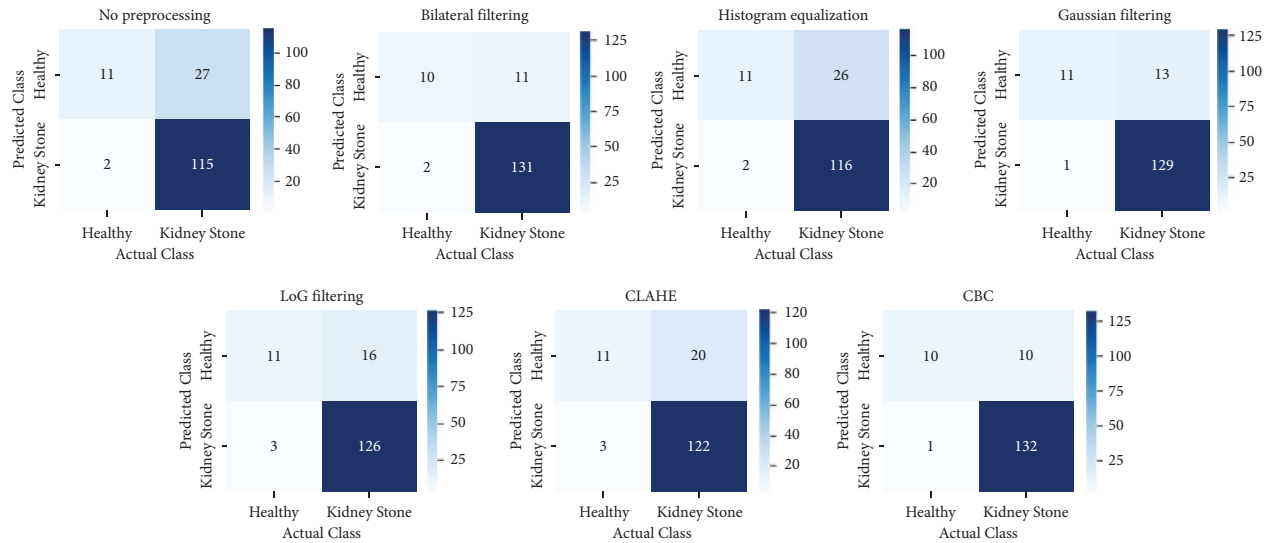
FIGURE 9: Confusion matrix for the YOLOv4 models.

FIGURE 10: Confusion matrix for the Mask R-CNN models.

TABLE 5: The performance comparison of YOLOv4 and Mask R-CNN models at 0.5 IoU threshold.

Preprocessing	Model	Accuracy	Precision	Recall	F1-score	Specificity
No preprocessing	YOLOv4	83.2	98.3	83.1	90.1	84.6
No preprocessing	Mask R-CNN	81.8	99.1	80.9	89.1	91.7
Bilateral filtering	YOLOv4	94.1	98.5	95.1	96.8	81.8
Bilateral filtering	Mask R-CNN	91.6	98.5	92.2	95.3	83.3
Histogram equalization	YOLOv4	83.9	97.6	84.5	90.6	76.9
Histogram equalization	Mask R-CNN	81.9	98.3	81.7	89.2	84.6
Gaussian filtering	YOLOv4	92.3	98.5	92.9	95.7	84.6
Gaussian filtering	Mask R-CNN	90.9	99.2	90.8	94.9	91.7
Laplacian of Gaussian	YOLOv4	88.5	96.9	90.1	93.4	73.3
Laplacian of Gaussian	Mask R-CNN	87.8	97.7	88.7	92.9	78.6
CLAHE	YOLOv4	87.3	96.9	88.7	92.6	73.3
CLAHE	Mask R-CNN	85.3	97.6	85.9	91.4	78.6
CBC	YOLOv4	**96.1**	**99.3**	**96.5**	**97.9**	**91.7**
CBC	Mask R-CNN	92.8	99.2	92.9	96.0	90.9

TABLE 6: The performance comparison of pretrained models.

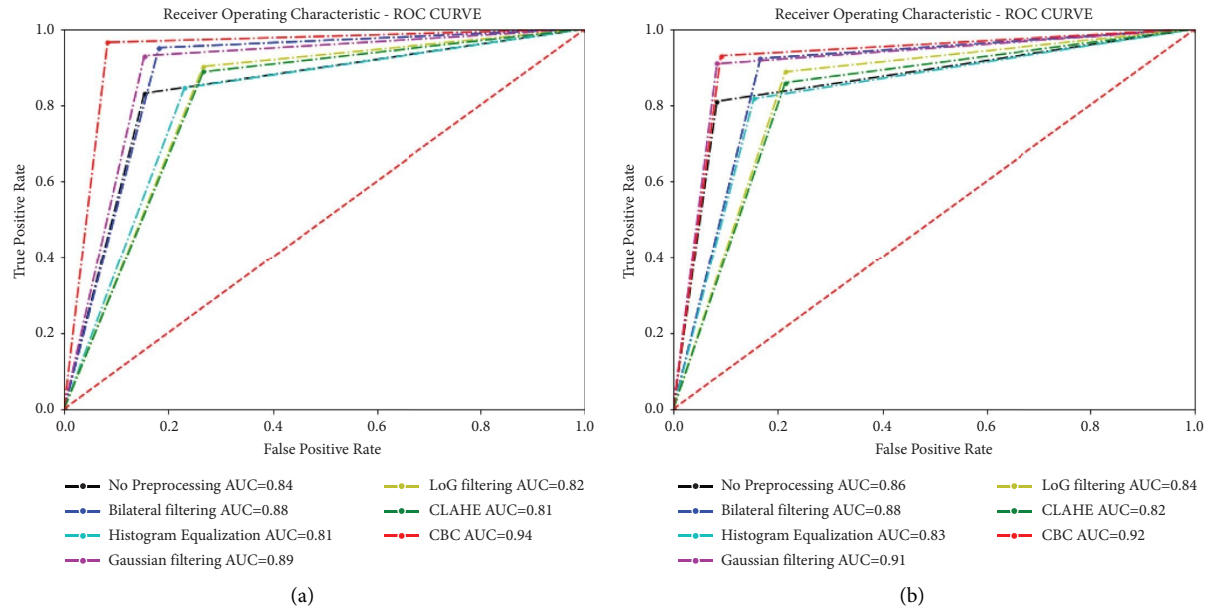| Pretrained model	Accuracy	Precision	Recall	F1-score
DenseNet201 [57]	89.3	89.5	99.6	94.3
EfficientNet-B0 [56]	87.3	89.3	97.3	93.1
MobileVnet2 [59]	86.0	89.0	96.1	92.4
ResNet101 [58]	87.8	89.3	97.9	93.4

FIGURE 11: (a) ROC of YOLOv4. (b) ROC of Mask R-CNN.

DenseNet201 [57] method, one of the transfer learning methods, is the most successful method with an accuracy of 89.3%.

The receiver operating characteristics (ROC) curve is an evaluation curve to check the performance of any classification model. The ROC curve is widely used to evaluate the performance of machine learning algorithms. It is efficient, especially in unbalanced datasets and it explains how well the model predicts. The ROC curve has a false positive rate (FPR) on the x-axis and true positive rate (TPR) on the y-axis. It facilitates the comparison of the accuracy of different models trained on the same dataset. Figure 11(a) shows the ROC curve of the YOLOv4 models, and Figure 11(b) shows the ROC curve of the Mask R-CNN models used for this study. Area under curve (AUC) refers to the area under the ROC, and it can be considered as a summary of model performance. In this curve, the larger area leads the more accurate model predictions. The ideal value for AUC is 1. When Figure 11 is examined, it is seen that the YOLOv4 model with CBC preprocessing step has the highest AUC (0.94). This means that the predictions of the proposed model are correct with a 94% probability.

5. Discussion and Conclusion

In this study, a CNN-based computer-aided diagnostic system was proposed to automatically detect kidney stones in DUSX images. For this purpose, a new dataset was proposed to the literature obtaining 630 DUSX images that belong to the patients of Ataturk University's Urology Department. We believe that the proposed dataset paves the way for further investigation of kidney stone detection systems.

The presence of noise in DUSX images and poor quality, especially in contrast form, reduces the success of CNN-based models in kidney stone detection. Six image enhancement techniques (GF, LoG, BF, HE, CLAHE, and CBC) were evaluated to ensure the quality of DUSX images. We investigated the effect of these techniques on automatic kidney stone detection. The experimental results show that the YOLOv4 model using the CBC technique as a preprocessing step has the best performance with 96.1% accuracy rate and this technique was proposed as a promising result model. The success of the proposed result model was clinically evaluated and accepted by a specialist urologist. The model can help urologists and radiologists to accurately detect kidney stone cases and reduce their workload. Additionally, we expect that the use of our proposed model will help to reduce the unnecessary radiation exposure and associated medical costs that come with CT scans.

We encountered some challenges during the training phase of the proposed system. Since YOLOv4 and Mask R-CNN architectures perform operations such as feature extraction and size reduction directly on the training images, the training times lasted too long. Moreover, labeling the images one by one before the training phase, especially the polygon tagging used

by the Mask R-CNN architecture, brings a serious workload to the operational processes. When we evaluate the YOLOv4 and Mask R-CNN architectures in terms of training time, the workload on operational processes, and detection performances, it is observed that YOLOv4 outperformed Mask R-CNN in terms of workload and training time. Another challenge is that patient or device movements can cause blurry or distorted images, making it difficult to detect kidney stones. Additionally, such actions lead to an increase in false positive results. Moreover, all DUSX images obtained from a single hospital may limit the generalizability of the model. In future studies, we aim to expand our dataset obtaining DUSX data from different hospitals to generalize the performance of our model and make it more robust.

In future studies, the dataset will be expanded and a balanced data distribution will be established to enhance the accuracy and precision of kidney stone detection from images. In addition, we have planned to achieve the detection of smaller kidney stone types with higher accuracy and speed by mapping the locations of the stones using segmentation methods. Moreover, we will evaluate the ability of our model to detect other pathological conditions such as tumors and cysts, and optimize the model for such situations.

Data Availability

The dataset used for this study is available at https://github.com/ugrkilc/DUSX-Dataset. The code that supports the findings of this study is available from the first author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

In this article, all researchers made significant contributions. Ugur Kilic has the main responsibility of the study from the implementation of the methods to validation of the system. Isil Karabey Aksakalli has an intermediary role for obtaining the medical data and data curation, writing, reviewing and editing, and visualization roles. Gulsah Tumuklu Ozyer and Baris Ozyer are experienced in deep learning, and they contributed to combined image enhancement methods. Tugay Aksakalli and Senol Adanur extracted the DUSX images from the Ataturk University Research Hospital, and they analysed and tagged the images.

Acknowledgments

References

[1] J. Serrat, F. Lumbreras, F. Blanco, M. Valiente, and M. López-Mesas, "myStone: a system for automatic kidney stone classification," *Expert Systems with Applications*, vol. 89, pp. 41–51, 2017.

[2] B. Medical, "Medical gallery of blausen medical 2014," *WikiJournal of Medicine*, vol. 1, no. 2, pp. 1–79, 2014.

[3] M. R. Devi and S. Suganyadevi, "DE-NOISING of kidney stone ultrasound image using prewittsmooth convolution filter (PSCF)," *Journal or Critical Reviews*, vol. 7, no. 10, pp. 1779–1787, 2020.

[4] I. K. Aksakalli, S. Kaçdioğlu, and S. Hanay, "Kidney X-ray images classification using machine learning and deep learning methods," *Balkan Journal of Electrical and Computer Engineering*, vol. 9, no. 2, pp. 144–151, 2021.

[5] C. Türk, A. Petřík, K. Sarica et al., "EAU guidelines on diagnosis and conservative management of urolithiasis," *European Urology*, vol. 69, no. 3, pp. 468–474, 2016.

[6] M. Gottlieb, B. Long, and A. Koyfman, "The evaluation and management of urolithiasis in the ED: a review of the literature," *American Journal of Emergency Medicine*, vol. 36, no. 4, pp. 699–706, 2018.

[7] G. C. Colleran, M. J. Callahan, H. J. Paltiel et al., "Imaging in the diagnosis of pediatric urolithiasis," *Pediatric Radiology*, vol. 47, no. 1, pp. 5–16, 2017.

[8] J. Gao, Q. Jiang, B. Zhou, and D. Chen, "Convolutional neural networks for computer-aided detection or diagnosis in medical image analysis: an overview," *Mathematical Biosciences and Engineering*, vol. 16, no. 6, pp. 6536–6561, 2019.

[9] M. Längkvist, J. Jendeberg, P. Thunberg, A. Loutfi, and M. Lidén, "Computer aided detection of ureteral stones in thin slice computed tomography volumes using Convolutional Neural Networks," *Computers in Biology and Medicine*, vol. 97, pp. 153–160, 2018.

[10] M. Puttagunta and S. Ravi, "Medical image analysis based on deep learning approach," *Multimedia Tools and Applications*, vol. 80, no. 16, pp. 24365–24398, 2021.

[11] X. Yao, X. Wang, S. H. Wang, and Y. D. Zhang, "A comprehensive survey on convolutional neural network in medical image analysis," *Multimedia Tools and Applications*, vol. 81, no. 29, pp. 41361–41405, 2022.

[12] A. Soni and A. Rai, "Kidney stone recognition and extraction using directional emboss and SVM from computed tomography images," in *Proceedings of the 2020 Third International Conference on Multimedia Processing, Communication and Information Technology (MPCIT)*, pp. 57–62, Shivamogga, India, December 2020.

[13] J. Verma, M. Nath, P. Tripathi, and K. Saini, "Analysis and identification of kidney stone using K th nearest neighbour (KNN) and support vector machine (SVM) classification techniques," *Pattern Recognition and Image Analysis*, vol. 27, no. 3, pp. 574–580, 2017.

[14] A. Khan, R. Das, and M. Parameshwara, "Detection of kidney stone using digital image processing: a holistic approach," *Engineering Research Express*, vol. 4, no. 3, Article ID 035040, 2022.

[15] S. Selvarani and P. Rajendran, "Detection of renal calculi in ultrasound image using meta-heuristic support vector machine," *Journal of Medical Systems*, vol. 43, no. 9, pp. 300–309, 2019.

[16] S. Eskandari, S. Meshgini, and A. Farzamnia, *Using a Novel Algorithm in Ultrasound Images to Detect Renal Stones*, Springer, Berlin, Germany, 2022.

[17] P. Chak, P. Navadiya, B. Parikh, and K. C. Pathak, *Neural Network and Svm Based Kidney Stone Based Medical Image Classification*, Springer, Berlin, Germany, 2019.

[18] A. Parakh, H. Lee, J. H. Lee, B. H. Eisner, D. V. Sahani, and S. Do, "Urinary stone detection on CT images using deep convolutional neural networks: evaluation of model performance and generalization," *Radiology Artificial intelligence*, vol. 1, no. 4, Article ID e180066, 2019.

[19] Y. Cui, Z. Sun, S. Ma et al., "Automatic detection and scoring of kidney stones on noncontrast CT images using STONE nephrolithometry: combined deep learning and thresholding methods," *Molecular Imaging and Biology*, vol. 23, no. 3, pp. 436–445, 2021.

[20] K. Yildirim, P. G. Bozdag, M. Talo, O. Yildirim, M. Karabatak, and U. R. Acharya, "Deep learning model for automated kidney stone detection using coronal CT images," *Computers in Biology and Medicine*, vol. 135, Article ID 104569, 2021.

[21] M. N. Islam, M. Hasan, M. Hossain, M. G. R. Alam, M. Z. Uddin, and A. Soylu, "Vision transformer and explainable transfer learning models for auto detection of kidney cyst, stone and tumor from CT-radiography," *Scientific Reports*, vol. 12, no. 1, pp. 11440–11514, 2022.

[22] M. Baygin, O. Yaman, P. D. Barua, S. Dogan, T. Tuncer, and U. R. Acharya, "Exemplar Darknet19 feature generation technique for automated kidney stone detection with coronal CT images," *Artificial Intelligence in Medicine*, vol. 127, Article ID 102274, 2022.

[23] K. K. Patro, J. P. Allam, B. C. Neelapu et al., "Application of Kronecker convolutions in deep learning technique for automated detection of kidney stones with coronal CT images," *Information Sciences*, vol. 640, Article ID 119005, 2023.

[24] Ö Sabuncu, B. Bilgehan, E. Kneebone, and O. Mirzaei, "Effective deep learning classification for kidney stone using axial computed tomography (CT) images," *Biomedical Engineering/Biomedizinische Technik*, vol. 68, no. 5, pp. 481–491, 2023.

[25] M. Akshaya, R. Nithushaa, N. S. M. Raja, and S. Padmapriya, "Kidney stone detection using neural networks," in *Proceedings of the 2020 International Conference on System, Computation, Automation and Networking (ICSCAN)*, pp. 1–4, Pondicherry, India, July 2020.

[26] H. El-Sheikh, O. Galal, and H. Ali, "Magnetic resonance imaging for detection of kidney stones in correlation with computed tomography," *Benha Journal of Applied Sciences*, vol. 7, no. 5, pp. 107–115, 2022.

[27] S. S. Parvathi, B. S. Chandana, and J. Harikiran, "Depth invariant 3D-CU-net model with completely connected dense skip networks for MRI kidney tumor segmentation," *Traitement du Signal*, vol. 40, no. 1, pp. 217–225, 2023.

[28] M. Kobayashi, J. Ishioka, Y. Matsuoka et al., "Computer-aided diagnosis with a convolutional neural network algorithm for automated detection of urinary tract stones on plain X-ray," *BMC Urology*, vol. 21, no. 1, pp. 102–110, 2021.

[29] W. Preedanan, K. Suzuki, T. Kondo et al., "Urinary stones segmentation in abdominal X-ray images using cascaded U-net pipeline with stone-embedding augmentation and lesion-size reweighting approach," *IEEE Access*, vol. 11, pp. 25702–25712, 2023.

[30] K. Viswanath and R. Gunasundari, "Design and analysis performance of kidney stone detection from ultrasound image by level set segmentation and ANN classification," in *Proceedings of the 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 407–414, Delhi, India, September 2014.

[31] A. Dutta, A. Gupta, and A. Zissermann, "VGG image annotator (VIA)," 2016, https://www.robots.ox.ac.uk/%7Evgg/software/via/.

[32] D. Nandan, J. Kanungo, and A. Mahajan, "An error-efficient Gaussian filter for image processing by using the expanded operand decomposition logarithm multiplication," *Journal of Ambient Intelligence and Humanized Computing*, vol. 2018, pp. 1–8, 2018.

[33] A. S. Mohamad, N. S. A. Halim, M. N. Nordin, R. Hamzah, and J. Sathar, "Automated detection of human RBC in diagnosing sickle cell anemia with Laplacian of Gaussian filter," in *Proceedings of the 2018 IEEE Conference on Systems, Process and Control (ICSPC)*, pp. 214–217, Melaka, Malaysia, December 2018.

[34] R. C. Gonzalez, *Digital Image Processing*, Pearson education india, Hoboken, NJ, USA, 2009.

[35] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proceeding of the 6th International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, pp. 839–846, Bombay, India, January 1998.

[36] N. H. Kaplan and I. Erer, "Bilateral pyramid based pan-sharpening of multispectral satellite images," in *Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium*, pp. 2376–2379, Munich, Germany, July 2012.

[37] S. Paris, "A gentle introduction to bilateral filtering and its applications," in *Proceedings of the ACM SIGGRAPH 2007 Courses*, San Diego, CA, USA, August 2007.

[38] S. M. Aktürk, *Grabcut etkileşimli bölütleme yöntemi üzerinde iyileştirme çalışmaları*, Ph.D. thesis, Karadeniz Teknik Üniversitesi, Trabzon, Türkiye, 2018.

[39] H. Yoon, Y. Han, and H. Hahn, "Image contrast enhancement based sub-histogram equalization technique without over-equalization noise," *World Academy of Science, Engineering and Technology*, vol. 50, pp. 176–182, 2009.

[40] T. Thamizhvani, K. Ahmed, R. Hemalatha, A. Dhivya, and R. Chandrasekaran, "Enhancement of MRI images of hamstring avulsion injury using histogram based techniques," *Multimedia Tools and Applications*, vol. 80, no. 8, pp. 12117–12134, 2021.

[41] K. Burçin and V. V. Nabiyev, "Dijital mamografi görüntülerinin kontrast sınırlı adaptif histogram eşitleme ile iyileştirilmesi," in *Proceedings of the VII. Ulusal Tıp Bilişimi Kongresi*, pp. 14–17, Gazimağusa, Turkey, October 2010.

[42] B. Türker, *Mamografi görüntülerinde matematiksel morfolojik filtreleme ile gürültü giderme ve kontrast iyileştirme*, Ph.D. thesis, Fen Bilimleri Enstitüsü, Ankara, Türkiye, 2020.

[43] S. M. Pizer, E. P. Amburn, J. D. Austin et al., "Adaptive histogram equalization and its variations," *Computer Vision, Graphics, and Image Processing*, vol. 39, no. 3, pp. 355–368, 1987.

[44] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only Look once: unified, real-time object detection," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, Las Vegas, NV, USA, June 2016.

[45] J. Redmon and A. Farhadi, "Yolov3: An Incremental Improvement," 2018, https://arxiv.org/abs/1804.02767.

[46] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "Yolov4: Optimal Speed and Accuracy of Object Detection," 2020, https://arxiv.org/abs/2004.10934.

[47] C. Y. Wang, H. Y. M. Liao, Y. H. Wu, P. Y. Chen, J. W. Hsieh, and I. H. Yeh, "CSPNet: A New Backbone that Can Enhance Learning Capability of CNN," 2020, https://arxiv.org/abs/1911.11929.

[48] B. Kilic, E. Baykal, M. Ekinci, H. Dogan, M. E. Ercin, and S. Ersoz, "Automated Nuclei Detection on Pleural Effusion Cytopathology Images Using YOLOv3," in *Proceedings of the*

2019 4th International Conference on Computer Science and Engineering (UBMK), Samsun, Turkey, September 2019.

[49] M. F. İnkaya, "Derin öğrenme algoritmaları kullanarak güvenli çocuk oyun alanlarında karşılaştırmalı tehlikeli nesne tespiti yapılması," M.Sc. thesis, Bursa Teknik Üniversitesi, Bursa, Türkiye, 2021.

[50] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-Cnn," 2017, https://arxiv.org/abs/1703.06870.

[51] A. N. Ömeroğlu, N. Kumbasar, E. A. Oral, and I. Y. Ozbek, "Mask R-CNN algoritması ile hangar tespiti hangar detection with mask R-CNN algorithm," in *Proceedings of the 2019 27th Signal Processing and Communications Applications Conference (SIU)*, pp. 1–4, Sivas, Turkey, April 2019.

[52] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-Cnn: Towards Real-Time Object Detection with Region Proposal Networks," 2015, https://arxiv.org/abs/1506.01497.

[53] R. Anantharaman, M. Velazquez, and Y. Lee, "Utilizing mask R-CNN for detection and segmentation of oral diseases," in *Proceedings of the 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 2197–2204, Madrid, Spain, December 2018.

[54] W. Abdulla, "Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow," 2017, https://github.com/matterport/Mask_RCNN.

[55] T. Y. Lin, M. Maire, and S. Belongie, *Microsoft Coco: Common Objects in Context*, Springer, Berlin, Germany, 2014.

[56] M. Tan and Q. Le, "Efficientnet: rethinking model scaling for convolutional neural networks," pp. 6105–6114, 2019, https://arxiv.org/abs/1905.11946.

[57] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," pp. 4700–4708, 2017, https://arxiv.org/abs/1608.0699.

[58] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," pp. 770–778, 2016, https://arxiv.org/abs/1512.03385.

[59] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "Mobilenetv2: inverted residuals and linear bottlenecks," pp. 4510–4520, 2018, https://arxiv.org/abs/1801.04381.