KIDNEY CANCEROUS TUMOR PREDICTION USING CNN SYSTEM ARCHITECTURE

M.F. Shahoriar Titu[1*], H.M. Emon[1],
S.A. Aumi[1], M.S. Bhuiyan[1],
M.F. Murshid[1], M.R. Rahman[1]

[1]Electrical and Computer Engineering.
North South University, Dhaka, Bangladesh.

*Corresponding author(s). E-mail(s): fahim.shahoriar@northsouth.edu

Article History: Received March 2, 2024; Revised May 20, 2024; Accepted May 23, 2024

ABSTRACT: The study addresses the challenge of interobserver variability in the treatment decisions for kidney cancers, a concern highlighted by the anticipated 73,750 kidney cancer diagnoses in the United States in 2020. This variability often arises due to the subtle differences in imaging characteristics of tumor subtypes. To address this issue, we propose an end-to-end deep learning model leveraging multi-phase CT scans to differentiate between five primary histologic subtypes of kidney cancers, encompassing both benign and malignant tumors. The proposed model demonstrates remarkable precision in identifying kidney cancers, even those of minimal size. In preparing the data for analysis, we divided it into training and validation test sets. The training set was used to employ the random forest method for ranking potential predictors based on their predictive importance. The model's performance was then validated on the test set using leave-one-out cross-validation. This study utilized convolutional and recurrent neural networks to predict kidney cancer outcomes. We used the models to classify adenoma, adenocarcinoma, and non-neoplastic whole slide images (WSIs). The evaluation of our models was conducted using three distinct test sets. The results showed area under the curve scores of 0.97 and 0.99 for distinguishing between cancerous tumors and adenomas and 0.96 and 0.99 for differentiating between kidney cancer and adenomas, respectively. These findings suggest that our models are not only generalizable but also hold significant potential to integrate and deploy into realistic pathological diagnostic workflows of kidney cancer.

KEYWORDS: *Convolution Neural Network, Recurrent Neural Network, Annotations, Augmentations.*

1.0 INTRODUCTION

Kidney disease has affected more than 10% of people worldwide, killing an immense number of people reliably, and innumerable people go through dialysis to stay aware of their lives [1]. Most patients also suffer from fever, erupting, and kidney sufferings, so this disease unfavorably influence the patient's physiology. The objective of therapy is to differentiate minimal cystic renal cell carcinoma from renal injuries, which affects an array of potential treatment plans for patients [2]. Hence, creating a reliable and efficient approach for detecting renal alterations is crucial for patient management. In the medical field, diagnoses of kidney diseases often employ X-rays, CT scans, and B-ultrasounds. CT scans target a specific body area to obtain a cross-sectional or tomographic image of the area under scrutiny. This method can deliver accurate, three-dimensional data of the examined body part, making organs, structures, and injuries easily identifiable. It's ability to be layered is a significant advantage that offers the possibility to reveal more complex information on post-examination [3]. The idea of kidney cancerous tumor prediction employing deep learning techniques was initially proposed by Dong [4]. The authors proposed a non-invasive layer-by-layer planning computation to upgrade the significant development in light of the robust AI network. Profound learning estimations have, furthermore, been extensively used in clinical settings. Tune et al. [5] used the piece fleecy C-infers estimation and the better Grow Cut computation to section kidney pictures, and the customarily delivered seed names raised the division efficiency. Xiang et al. [6] used cortical models and nonuniform advisers to depict the kidney structure in CT pictures. Xiong et al. [7] implemented a disease division system considering flexible distributed level sets, which truly isolated kidney malignant growths in ultrasound pictures. Gao [8] united the level set into picture division to deal with images with unbalanced faint characteristics and achieved extraordinary division results. Hu and his colleagues [9] carried the goalfollowing instrument into the ordinary picture division and achieved incredible division results. These results show the critical advantages of using significant learning-based instruments and work process structures to help clinical analyzers and histopathological characterization, especially concerning extended beginning of screening capability and diminishing logical twofold

examination. Another area for further improvement is the drawn-out course of social occasions, a significant grouping of WSIs with precise pixel-by-pixel remarks, which makes it difficult to set up a managed learning classifier with an alternate plan of pictures. This strategy would require critical development in the number of WSIs and taking care of resources. Based on data from 45,642 WSIs, Campanella et al. [12] employed this technique and obtained terrific results. On a dataset of 45,542 WSIs, significant learning has recently been used to perceive WSIs in kidney and kidney cancer; however, it is on a restricted scale. Sharma et al. [13] endeavored kidney harmful development requests utilizing a [8] WSIs dataset. Profound learning has been used in kidney illness to anticipate general perseverance [14], recognize centres [15], and plan development [7]. In this work, various deep learning models have been applied to portray kidney and kidney cancer-threatening development to assist cautious specialists' standard histology discoveries. We combined two datasets of kidney and kidney growth WSIs, each containing 4,238 and 4,128 WSIs, respectively [16]. This study shows that pre-trained models achieved better exactness and approval exactness. The MobileNetv2 model achieves close to 100% exactness and 96% precision. VGG16 and InceptionV3 gave 98% and 97% exactness individually and accomplished a precision of 96% and 97% separately. ResNet50 gave somewhat less exactness and approval precision, which are 98.8% and 95% apiece. The proposed CNN model got 98% exactness and 97% approval precision. The pre-prepared and custom CNN models gave higher exactness and approval exactness than the past investigations.

2.0 MANUSCRIPT PREPARATION

The dataset employed in this work has been gathered from various reliable opensource destinations and afterward consolidated to make an extensive dataset. The dataset comprises CT pictures of kidney disease positive patients and typical patients. In the applied CNN model, three Conv2D layers, three MaxPooling2D layers, one smoothed layer, two thick layers, and a redressed direct unit enactment capability have been incorporated. The enactment capability utilized for the last thick layer was SoftMax. VGG16, MobileNetV2, InceptionV3, and ResNet50 were utilized for preprepared models, with minor changes in the last layers and a head model obtained from the essential model. Normal pooling, leveling, thick,

and dropout are the adaptable last layers. The CNN model is appropriate for highlight extraction since it removes qualities from provided pictures and learns and recognizes pictures in light of these elements. The intricacy of the learned attributes develops with each layer.

2.1 Materials and Tools

Python libraries are utilized to make AI, information science, information perception, picture and information control, and numerous applications. Given Python's broad library access, profound learning-based issues are especially effective with Python programming. Google Colab was utilized to deal with the vast datasets and prepare the model.

2.2 Dataset Description

The dataset utilized in this study consists of kidney CT pictures of 2,212 images [16]. The dataset has four classes, which are partitioned into an 8:1:1 ratio. The CT images of a Kidney Cancer patient and a normal patient are shown in Figure 1.

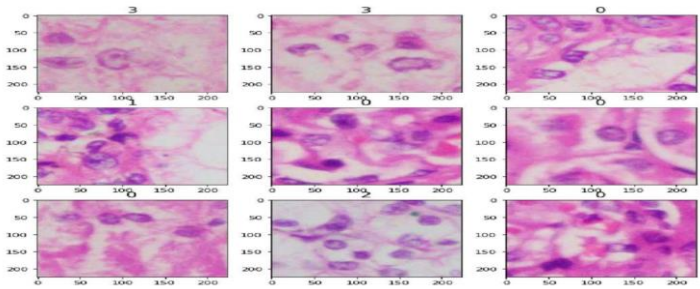

Figure 1: Sample CT images of the employed dataset

Next, the training images was fed traditional preprocessing techniques before fitting into the applied deep learning models, which included bringing in pictures of a particular size, separating the dataset, and using information expansion methodology. The precision was further developed once the model was fitted and hyperparameters were tweaked, employing the validation subset of images.

2.3 System Architecture

There are three channels in the input layer, and the info shape is 224×224. The channel size is 32 with cushioning, the portion size is 3, and the

enactment capability is ReLU in the principal layer of the proposed engineering. The main max- pooling layer follows, which has a pool size of 2 and steps of 1. The following layer is a level layer that joins each of the pooled highlights into one segment. Two thick layers were created eventually. The initiation capability for the primary layer is ReLU, though the enactment capability for the most un-thick layer is SoftMax. Figure 2 illustrates the architecture of the proposed CNN model.

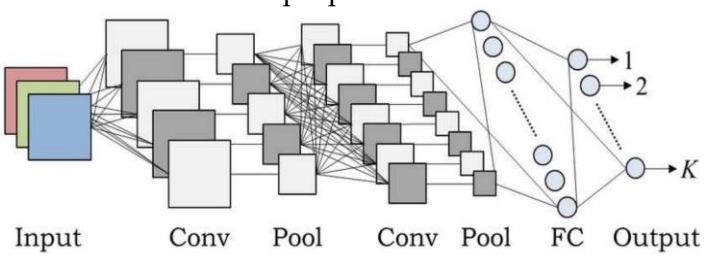

Figure 2: CNN System architecture

2.3.1 Convolutional Layer

The convolutional layer is the essential layer of the CNN. On CNN, it is the center structure block. Most of the calculations happens in this layer. Input information, a channel, and an included application are expected for this layer. In this layer, the information pictures have gone through a channel. The result of similar channels gets the capability map from the convolutional layer. In picture arrangement errands, at least one 2D lattices are viewed as the contribution to the convolutional layer and different 2D frameworks are started as the result. Information and result network numbers might be isolated [17]. The activations of the convolutional layer can be computed by using the following equation:

$$C_{ij} = \theta\left(\sum_{n=1}^{m} W_{n_1 s_1} P_{nt+1-t} + b_i\right) \quad (1)$$

where C_{ij} represents the output of the convention layer for the i^{th} feature map of the j^{th} band, m is the filter scale, W_n, i is the weight vector for the n^{th} band of the i^{th} filter, b_i denotes the bias for the i^{th} feature map, and Θ is the activation function.

2.3.2 Pooling Layer

The pooling layer is another fundamental structure block of CNN. In CNN's layers, the two most normal pooling frameworks are max pooling and normal pooling. The pooling layer eliminates the repetitive elements from the picture and makes the picture all around informed. While utilizing max pooling, the layer takes the limit of a locale from the channel's ongoing perspective each time and assists with the main elements of the picture. The normal pooling layer implies the worth of the ongoing perspective each time. A pooling layer is fundamental for reducing the quantity of element guides and organization boundaries, and a dropout layer forestalls overfitting. The activation of max pooling can be computed as:

$$P_{ij} = max_{k=1}^{t}\left(h_{i,(j-1)}(l+g)\right) \quad (2)$$

In (2), P_{ij} is determined by the i^{th} function map and the j^{th} band of the pooling layer. The pooling scale, represented by t, refers to the total number of bands that are pooled together. Additionally, the subsampling factor, symbolized by l , plays a crucial role in this process

2.3.3 Flattened Layer

After a grouping of convolutional and pooling procedure on the highlights portrayal of the pictures, the Straightened layer is utilized in the result pictures into a solitary long consistent direct exhibit or a vector. The most common way of changing over all the resultant 2D exhibits into a vector is called Leveling. A smoothed layer is ready for the following connected layer of picture grouping by changing over it into a one-layered cluster.

2.3.4 Fully Connected Layer

There are supports of completely associated layers after convolution and pooling layers in the CNN. CNN exceptionally relies upon completely associated layers. In PC vision picture acknowledgment and characterization, completely associated layers have been perceived as exceptionally valuable. The completely associated layer's bits of feedback function as results of the last pooling or convolutional layers, that are leveled and afterward kept in the completely associated layers. For the highest point of these learned elements, completely associated layers function as a categorizer. As a completely associated layer, the ReLU

initiation capability is generally utilized.

2.4 Pretrained Models

In this work, three CNN-based pre-prepared models were utilized, i.e., VGG16, MobileNetv2, and InceptionV3. The CT pictures are the initial step, and the stacking of a pre-prepared model is the second. Three preprepared models are stacked in the subsequent segment. The stacked pre-prepared models were changed involving the accompanying layers in the third part. Finally, the outcome will be introduced as Kidney Sickness contaminated and typical patients in the result area. The convolutional brain network Inceptionv3 has a profundity of 50 layers. With the ImageNet loads, the preprepared form of Inceptionv3 can sort up to 1,000 items. The ResNet50 model was additionally prepared utilizing the ImageNet dataset [19]. MobileNetV2 builds the state-of-the- art execution of adaptable models over a scope of model sizes on various tasks and seat stamps. In each line, MobileNetV2 works as a progression of n rehashing layers [20]. A basic portion can successfully remove the qualities from CT pictures [18]. VGG16 was utilized in this review and a fitting layer was added for the end product.

3.0 Results And Analysis

To conclude, five models had been trained using a trained generator and a validation generator. To train this model, a training and validation dataset was used. Train the custom CNN model and get 98% accuracy and 97% validation accuracy in the final epoch. The total number of epochs was 10. The model achieved satisfactory accuracy from the beginning. In the first epoch, it obtained 86% accuracy and 95% validation accuracy. The pre-trained model VGG16 has 98% accuracy and 96% validation accuracy in the 10th epoch; the training loss and validation loss were 3% and 8%, respectively.

Table 1: Performance Metrics of The Applied Models

Model	Accuracy (%)	Validation Acc (%)	Loss (%)	Validation Loss (%)
Custom CNN	98	97	3	8
MobileNetV2	99	96	14	62
VGG16	98	96	3	8
ResNet50	93	95	20	24
InceptionV3	97	97	32	20

MobileNetV2 secured the highest accuracy among the applied models, which is 99% accuracy and 96% validation accuracy. InceptionV3 has 97% accuracy and 97% validation accuracy. In InceptionV3, 32% training loss and 20% validation loss were observed. The histories of accuracy, validation accuracy, loss, and validation loss of the five models are given in Table 1.

Throughout this research, transfer learning was also used with four pre-trained models. The pre-trained models made much smoother predictions. In the VGG16 model, the accuracy was 98% and the validation accuracy was 96%. In the first epoch of the model, it obtained 92% accuracy and 95% validation accuracy. It gradually increased and obtained 98% accuracy in the 10th epoch. For validation accuracy, it secured 95% in the first epoch, increased to 99% at the 6th epoch, and gradually decreased to 96% at the 10th epoch. The model provided 3% training loss and 8% validation loss. In the first epoch, it provided a 20% training loss, which gradually decreased to 3% in the 10th epoch. For the validation loss, it provided 8% in the first epoch and also in the 10th epoch. In the middle epochs, it gradually decreased. MobileNetV2 provided the highest accuracy among these pre-trained models. It obtained 99% accuracy and 96% validation accuracy. Compared to the VGG16 model's accuracy, it got higher accuracy in the first epoch, which is 94%. Although their validation accuracy for the first epoch was the same, which is 95% for both models, MobileNetV2 has a 14% training loss and a 62% validation loss. In this case, MobileNetV2 did not provide a good validation loss. MobileNetV2's validation loss is 62%, which is the highest validation loss among these four models. The confusion matrix of the custom CNN model is shown in Figure 3.

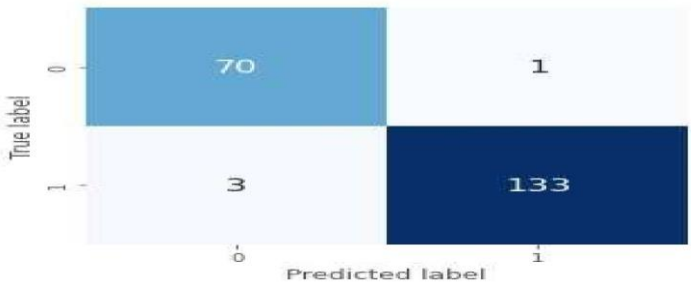

Figure 3: Confusion Matrix of the applied custom CNN model

Predictions, data, and features are the three major terms in error analysis. Error analysis based on prediction can be executed using a confusion matrix, where the percentage can be divided into false positives, false negatives, true positives, and true negatives. The size and nature of the data are also significant for error analysis. The precision of a dataset is defined as the ratio of accurate predictions to provide the total predictions. It can provide us with a brief overview of a model's training quality and potential performance. It does not, however, offer thorough instructions on how to apply it to the issue. When the expense of false positives is substantial, precision—also referred to as PPV—is an appropriate metric to employ. When the cost of false negatives is high, the optimal model is selected using the recall model metric. When the cost of false negatives is significant, recall becomes valuable. When attempting to assess how well recall and accuracy are balanced, an F1-score is required. It is a broad indicator of the correctness of the model. It integrates recollection and accuracy. Low false positives and false negatives are the keys to a high F1-score. During the model evaluation, the test dataset was utilized. This test dataset was created during the dataset split and was reserved solely for testing purposes. This study calculated the precision, recall, and F1-score for all the models utilizing the test dataset. Table 2 demonstrates the precision, recall, and F1-score of the applied VGG16, InceptionV3, MobileNetV2, ResNet50, and Custom CNN models for the proposed kidney cancerous tumor prediction system. It shows that VGG16 and InceptionV3 have the highest F1-score among all these models.

Table 2: Precision, Recall And F1 Score Of The Applied Models

Model	**State**	**Precision**	**Recall**	**F1**
Custom CNN	kidney cancer	0.89	0.97	0.93
Custom CNN	Normal	0.97	0.88	0.93
VGG16	kidney cancer	0.95	0.92	0.94
VGG16	Normal	0.92	0.96	0.94
InceptionV3	kidney cancer	0.96	0.89	0.93
InceptionV3	Normal	0.9	0.96	0.93
MobileNetV2	kidney cancer	0.97	0.88	0.92
MobileNetV2	Normal	0.89	0.97	0.93
ResNet50	kidney cancer	0.87	0.98	0.92
ResNet50	Normal	0.98	0.86	0.91

In this study, testing was conducted by feeding real-time CT images to the trained model. Once training had finished, the model was saved, and separate HDF5 files were generated for each of the four different models. These four files were used to make the predictions. Each model was fed an individual CT image as input, and following the input, the model predicted whether the image indicated kidney cancer or a normal CT image. Figure 4 illustrates the prediction result, indicating whether it belonged to a kidney cancer patient or a normal individual.

In Figure 4, a CT image affected by a kidney cyst was used as an input to the model. The model identified the CT image as a kidney cancer, which was predicted precisely.

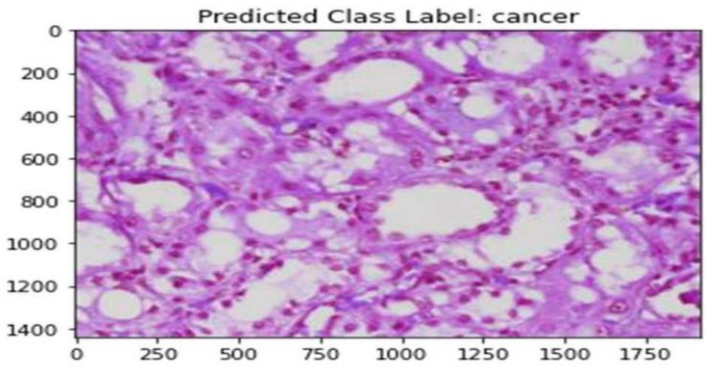

Figure 4: Sample CT image of the Kidney cancer affected result

Following the test, another image was fed as input, and the model correctly predicted it as a tumor-affected CT image. Figure 5 depicts the model's predictions of a tumor-affected CT image.

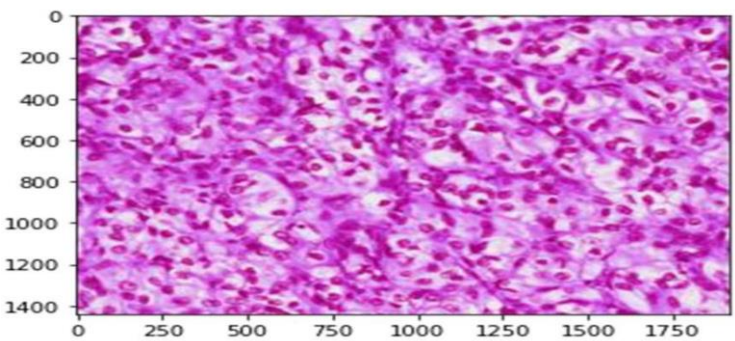

Figure 5: Sample CT image of the Normal result

Table 3 portrays that all the models had an excellent result. Pre-trained models demonstrated higher accuracy compared to the reference models. Comparatively inceptionV3, VGG16 and the custom CNN model had achieved the highest accuracy than the previously studied models. Mobile netV2 had attained the highest accuracy at 99%. VGG16 and InceptionV3 had achieved the level of accuracy respectively 98% & 97%. Which surpassed the accuracies reported in the referenced article (95.9% for VGG16 and 93% for inceptionV3). ResNet50 had achieved excellent accuracy level of 93.8%, Surpassing the performance in previous studies where it achieved 88% [14] and 92.6% [13] accuracy.

Table 3 Comparison Of This Work With Other Similar Systems

Reference	Model Name	Acc (%)	In this study (%)
[10]	VGG16	95.9	98
[12]	MobileNetV2	97.4	99
[11]	InceptionV3	93	97
[14]	ResNet50	88	93.8
[13]	ResNet50	92.6	93.8
[4]	Custom CNN	93	98
[5]	Custom CNN	94.5	98

4.0 CONCLUSION

This research proposes a deep learning technique for CT image-based kidney cancer diagnosis. We employed four pre-trained models and a bespoke CNN model to distinguish between benign and malignant kidney cancers. We demonstrated on a dataset of 2212 CT scans that our models outperformed the earlier techniques. Additionally, we provide some examples of confusion matrices and forecasts. Furthermore, we have identified the difficulties and consequences of our models for future research and clinical practice. We have suggested specific methods to enhance the data quality, incorporate more clinical details, and to expand the application of our models to different cancer types. Our research has attained excellent dependability by utilizing many popular deep-learning models and it has the potential to be the foundation for a system that aids physicians in using CT scans to identify kidney cancer. Future research could improve this system using this method, which might be enhanced by creating a web application and utilizing a larger dataset. This would raise the system's efficacy, accuracy, and outcomes. With further research and development kidney cancer will be diagnosed and identified at the earlier stage. Thus, the patient will get the essential therapy regarding the condition of the kidney cancer. It will help them to cure the cancer earlier and lead them to the sound life.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare and are in agreement with the contents of the manuscript.

ACKNOWLEDGMENTS

We extend our heartfelt gratitude to our collaborators for their invaluable support throughout this research.

5.0 REFERENCES

[1] R. L. Chevalier, Evolutionary nephrology," *Kidney International Reports*, vol. 2, pp. 302–317, 2017.

[2] J. Eichberger, Kidney symptoms start about five days after exposure, Johns Hopkins study finds — Hub,'' 2020.

[3] Y. M. Bar-On, A. Flamholz, R. Phillips, and R. Milo, Sars-cov-2 (Kidney cancer) by the numbers," *eLife*, vol. 9, 2020.

[4] L. Wang, Z. Q. Lin, and A. Wong, TUMOR-Net: A Tailored Deep Convolutional Neural Network Design for Detection of Kidney Cancer Cases from Kidney X-Ray Images,'' *Scientific Reports*, vol. 10, pp. 1– 12, 2020.

[5] M. Heidari, S. Mirniaharikandehei, A. Z. Khuzani, G. Danala, Y. Qiu, and B. Zheng, Improving the performance of CNN to predict the likelihood of Kidney Cancer using kidney X-ray images with preprocessing algorithms," *International Journal of Medical Informatics*, vol. 144, pp. 104284, 2020.

[6] A. Waheed, M. Goyal, D. Gupta, A. Khanna, F. Al-Turjman, and P. R. Pinheiro, TumorGAN: Data Augmentation Using Auxiliary Classifier GAN for Improved Kidney cancer Detection,'' *IEEE Access*, vol. 8, pp. 91916–91923, 2021.

[7] Q. Li, Z. Yu, Y. Wang and H. Zheng, TumorGAN: Data Augmentation Using Auxiliary Classifier GAN for Improved Kidney cancer Detection," *Sensors*, vol. 20, 2020.

[8] N. Narayan Das, N. Kumar, M. Kaur, V. Kumar, and D. Singh, Automated Deep Transfer Learning-Based Approach for Detection of Kidney Cancer Infection in Kidney X-rays,'' *IRBM*, 2020.

[9] S. Ying et al., Deep learning Enables Accurate Diagnosis of Novel Kidney (Kidney Cancer) with CT images," *medRxiv*, 2020.

[10] A. Shelke et al., Kidney X-ray Classification Using Deep Learning for Automated Kidney Cancer Screening,'' *SN Computer Science*, vol. 2, pp. 1–9, 2021.

[11] L. Gaur, U. Bhatia, N. Z. Jhanjhi, G. Muhammad, and M. Masud, Medical imagebased detection of Kidney Cancer using Deep Convolution Neural Networks," *Multimedia Systems 2021*, vol. 1, pp. 1– 10, 2021.

[12] I. D. Apostolopoulos and T. A. Mpesiana, Kidney cancer: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks,'' *Physical and Engineering Sciences in Medicine*, vol. 43, 20202.

[13] A. M. Ismael and A. S¸engu¨r, Deep learning approaches for Kidney Cancer detection based on kidney X-ray images," *Expert Systems with Applications*, vol. 164, 2021.

[14] A. Uddin, B. Talukder, M. M. Khan, and A. Zaguia, Study on Convolutional Neural Network to Detect Kidney Cancer from Kidney X- Rays,'' *Mathematical Problems in Engineering*, vol. 2021, 2021.

[15] M. Alruwaili, A. Shehab, and S. Abd El-Ghany, Kidney Cancer Diagnosis Using an Enhanced Inception-ResNetV2 Deep Learning Model in X-RAY Images," *Journal of Healthcare Engineering*, vol. 2021, 2021.

[16] KIDNEY CANCER Radiography Database—Kaggle.'' https://www.kaggle.com/datasets/atreyamajumdar/kidney-cancer

[17] R. Nanculef, P. Radeva, and S. Balocco, Training Convolutional Nets to Detect˜ Calcified Plaque in IVUS Sequences," *Intravascular Ultrasound*, pp. 141–158, 2020.

[18] C. Sitaula and M. B. Hossain, Attention-based VGG-16 model for Kidney Cancer kidney X-ray image classification,'' *Applied Intelligence*, vol. 51, pp. 2850–2863, 2020.

[19] K. He, X. Zhang, S. Ren, and J. Sun, Deep Residual Learning for Image Recognition," *Conference on Computer Vision and Pattern Recognition*, vol. 2016, pp. 770– 778, 2015.

[20] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, MobileNetV2: Inverted Residuals and Linear Bottlenecks,'' *Conference on Computer Vision and Pattern Recognition*, pp. 4510–4520, 2018.