PREDICTION OF RENAL CELL CARCINOMA NEUROBLASTOMA USING MACHINE LEARNING

Muhammad Shoaib Kareem
Department of CS&IT
Riphah International University,
Lahore, Pakistan

Zeeshan Ahmad
Department of Electrical
Engineering,
Islamia University of Bahawalpur,
Pakistan

Talha Farooq Khan
Department of Computer Science,
Islamia University of Bahawalpur,
Pakistan

Mohsin Shahzad
Department of Electrical
Engineering,
Islamia University of Bahawalpur,
Pakistan

Mohsin Bashir
Department of Computer Science,
Riphah International University,
Lahore, Pakistan

Fawwad Hassan Jaskani
Department of Computer
Engineering, Islamia University of
Bahawalpur,
Pakistan

Fareeha Zahoor
Department of Electrical Engineering,
Islamia University of Bahawalpur,
Pakistan

Abstract— **Tumor Detection or Kidney Carcinoma Tumor is type of Kidney Cancer, which is very hazardous for health. There is a need of a proper analysis on this disease to treat the patients Kidney Tumor. Therefore, we have proposed an intelligent Tumor detection system. The main purpose of this research is to develop and implement an automated method that will help to detect and classify the Kidney Cancerous Tumor by processing Kidney images of Kits 19 dataset. We have used Kidney Tumor Segmentation dataset for testing our proposed methodology, it contains 300 CT-Scans images of Kidney Cancer patients. We have applied pre-processing techniques on these images such as morphological filtering, to enhance images. After obtaining the resultant image we have applied augmentation. We have used modified form of Convolutional Neural Network for classification purpose. The proposed technique provides a sophisticated diagnosis and 3 cross folds of prediction and validation loss with dice score and cross-entropy when compared with previous techniques. The parameter we used to validate the performance of our proposed technique is cross folding's, dice score and cross-entropy.**

Keywords— **Deep Learning, Computer Vision, Kidney Diseases**

I. INTRODUCTION

The chief clinical application of kidney imaging comprises excluding rescindable reasons of acute kidney grievance, like urinary obstruction, or recognizing permanent Chronic Kidney Disease (Kidney Carcinoma) that prevents unnecessary diagnosis like kidney biopsy [1]–[3]. Its non-invasiveness, little cost, lack of ionizing radioactivity, and wide accessibility make it a gorgeous option for recurrent monitoring and complement of the longitudinal variation in kidney dimension and sonographic features of kidney cortex related to kidney useful change. Though, the high individual variability in image gaining and clarification makes it problematic to interpret experience-based calculation into consistent practice, like invasive blood serum creatinine amount. Up till now, non-invasive imaging methods for organ practical and structural classification have been progressively investigated directing to diminish the invasive method in both problem-solving and screening situations. In [4] author has newly anticipated an inimitable pediatric Kidney Carcinoma maintenance model corresponding cost and minimizing incisiveness with the development of risk estimation by utilizing kidney CT imaging to predict the advance of Kidney Carcinoma and clinical conclusions in infants.

Conservatively, nephrologists lean towards kidney dimension, volume, cortical width and echogenicity to assess the sternness

of kidney injury. Identical short renal measurement (e.g., <8 cm), deceptive white cortex, limited capsule contour, entirely indicate a permanent kidney failing procedure with great specificity but restricted sensitivity [5]. Additionally, whether these scans parameters can be utilized to predict precise Estimated Glomerular Filtration Rates (eGFRs) remains argumentative. For example, [6], [7] studies have described that even though kidney dimension is highly detailed in detecting permanent Kidney Carcinoma, its association with eGFR was solitary week to modest, ranging no connotation to 0.66. Even, research using the conventional modification diet in renal disease research (MDRD) equation to evaluation eGFR (MDRD-eGFR) are measured, the finest correlation well-known between kidney dimension and eGFR has been 0.36. Likewise, a fair-to-moderate association of kidney dimensions and cortical echogenicity through eGFR has been stated. By disparity, cortical thickness looks to be healthier connected with MDRD-eGFR than the length of kidney, with a association coefficient as top as 0.85 [8]. Nevertheless, the research writing the correlation coefficient with 0.85 was attained for a minor sample of 42 grown person with Kidney Carcinoma without justification.

Yapark et al. established a Kidney Carcinoma scoring scheme integrating 3 CTS cans parameters, specifically kidney dimension, parenchymal width, and echogenicity, to recover the correlation; though, the association was modest (r = 0.587). [9] Also, the allotted score of each parameter remainders subjective with indeterminate inter-observer consistency.

To get rid of substantial inter-observer inconsistency in kidney CT Scans interpretation, Machine Learning (ML) provides a compact and objective basis for analytic calibration to notify clinical decisions. Current advances in imaging segmentation, registration, and classification through Deep Learning (DL) have significantly expanded the opportunity and scale of imaging analysis [10]. DL adapted to diagnostic applications may diminish unnecessary and offensive procedures, therefore greatly refining the sustainability and efficiency of current healthcare schemes.

Furthermore, with the extraordinary increase in computer science performance and real-time computer assisted diagnosis may additionally change telemedicine and mobile telecare. In the existing Kidney Carcinoma care prototypical, its relics controversial kidney role should be regularly partitioned in all symptom-less adults. The utmost commonly utilized Kidney Carcinoma screening assessments comprise testing the urine and blood for protein or serum creatinine; though, there is no decisive evidence signifying which screening trial is more suitable to the further in the situation of routine screening. Evolving an easily obtainable and non-invasive image indicator of kidney role using deep learning methods therefore may deliver a valuable courtesy tool for identifying Kidney Carcinoma. To discover this opportunity in clinical preparation, we established a deep learning algorithm which is based on kidney CT Scans imaging and medical data in a bulk registry-based Kidney Carcinoma cohort.
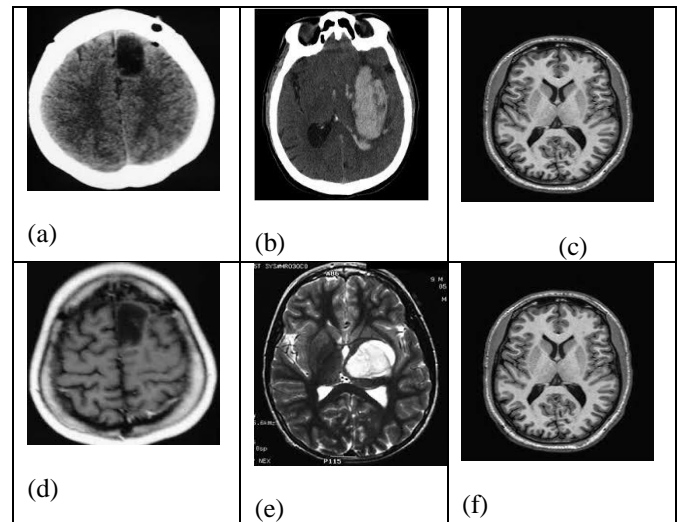

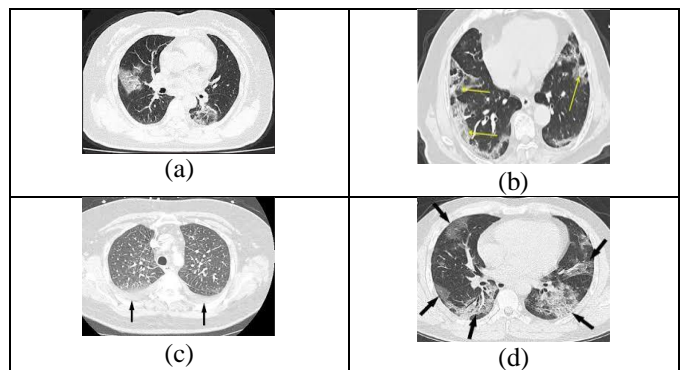*Figure 1 Normal kidney images*

Figure 2 Show some Diseases image of different test

In more than 3000 Kidney diseases affect 2 in every 3 people. It is very difficult to diagnose. 50 percent of the diagnoses in a public hospital is incorrect. The Kidney problem is extremely costly because of unnecessary visits, tests, and treatments to general practitioners because of incorrect diagnoses. Most people have also no access to a Haematologists. So, Machine Learning and Artificial Intelligence can be used to support image classification using Deep Neural Networks (DNN). When an image, data, or information of Kidney is input to a machine it recognizes the morphology of the Kidney disease and produces different diagnosis as the earlier prediction. To provide decision support this can be faster and better than the other traditional methods. ML and Deep Learning improve the knowledge and skills in primary care and substitutional detection in cost.

As there are two kinds of Kidney Tumors, malignant Tumor Detection and non-malignant. Malignant Tumor has a dangerous implication because it can extend and damage other portions of the body parts. In the case of the deadliest diseases

like Tumor, detection in early stages play a vital role in finding the probability of getting cured. At the primary stage, it is a challenge for a medical practitioner to differentiate a common Kidney tumor from modest benign tumors [8].

Kidney Cancer has affected the lives of many people and the detection of Kidney Cancer in initial stages plays an important role in determining the probability of getting survive. We can utilize modern Deep Learning techniques to facilitate early diagnosis and affordably inform individuals who may be at risk. This tool is not always accurate and is not intended to be a replacement for professional care. The manual patients' diseases detection of an image is a challenging and time concussing task due to different level of diseases. Hence, an automated and accurate system is required for patient diseases detection. The main characteristics and the contributions of this research are: Some pre-processing techniques are performed to create a high quality and a balanced image with test sets constructed only from images of tumors. For noise removal, we adapted the morphological filtering to shrink image regions and Conversion into Greyscale as it is difficult to operate an RGB image. We perform a filtering process named Black Hat, as it converts an image including the objects of the input image. We applied thresholding on images as it is a segmentation process. This is done to separate object pixels from background pixels. After that, we reconstruct our missing part of images using image Inpainting. To balance the imbalanced data, we have used augmentation techniques to increase the training dataset. We used Modified CNN model for the classification of our model.

II. PROPOSED ALGORITHM

We have used KiTS19 (Kidney Tumor Segmentation Dataset 2019) openly available on Kaggle, to achieve segmentation and classification of tumors. This set contains up to 1000 images. We will be creating high quality, a balanced dataset with test and validation sets built only from images of tumors with one associated image. This is important because if our model learns from an image of a given tumor, then it is tested on another image of the same tumor, our accuracy will be unrealistically inflated. A sample image taken from the dataset is shown in Fig below.
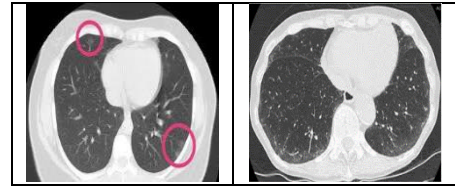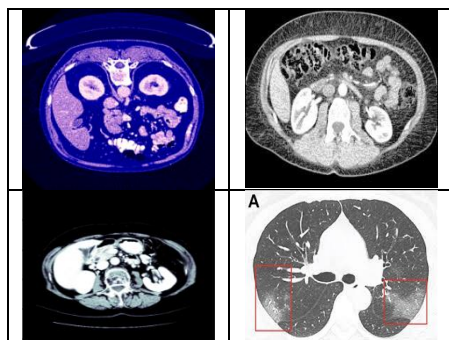

Figure 3 Patients Diseases sample image.

We have selected 300 kidney cancer patients CT scans images from dataset. Each pixel had to be labelled into one of three classes: Background, kidney, or tumor. The original scans have an image resolution of 512x512 and on average 216 slices (highest slice number is 1059).

We want to perform a 3-fold cross-validation for kidney tumor segmentation on the (KITS19) dataset. To achieve this, we are using our medical image segmentation techniques with Convolutional Neural Networks and Deep Learning. The aim of this research is to provide a fast building of medical image segmentation pipelines including data Input/Output, pre-processing, data augmentation, patch-wise analysis, metrics, a library with state-of-the-art deep learning models and model utilization like training, prediction as well as fully automatic evaluation (e.g., cross-validation). Our technique is based on Keras with Tensorflow as backend. We have selected 300 kidney cancer patients CT scans images from dataset. Each pixel had to be labelled into one of three classes: Background, kidney, or tumor. The original scans have an image resolution of 512x512 and on average 216 slices (highest slice number is 1059). A block diagram of our purposed methodology is shown in figure below
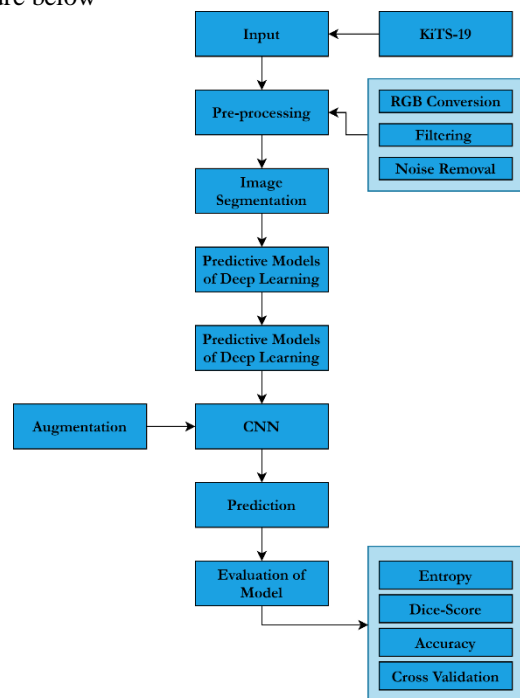

Figure 4 Block diagram of the proposed model.

In general, pre-processing techniques can handle the noise of the images and deep learning techniques learn to avoid these noises while making classification and predictions, though their ability to learn is conditioned upon the availability of large training data. In the case of medical Kidney images datasets, which are mostly imbalanced datasets, the Deep Learning models are easily susceptible to overfitting, i.e., they utilize visual hints such as noise indicators of the cancer category. The first step in the pipeline is to establish a Data I/O. It offers the utilization of custom Data I/O interfaces for fast integration of your specific data structure into the pipeline. We have used data pre-processing techniques to remove noise and noises on the images. We have modified the noise removal algorithm. The

The arrangement of sharpness, shadow, contrasts, and construction of the colour picture are novel procedures. To alter the Red Green Blue (RGB) into grayscale, an average method is used. The black hat is an image filtering process. We have used this filtering for the following reason. It proceeds an image that contains the basics of the input image that are lesser than the configuring component and dimmer than their surroundings. We have used the structuring element as a morphological operator. It affects the regions of images. We apply this filtering for eliminating the unnecessary regions in the images on the pixels of the image.

Binary Thresholding is a modest procedure of image segmentation. It is a technique to produce a binary image from an RGB image or grayscale. Typically, we use this to separate foreground pixels from the background pixels. In Inpaint methodology, we decompose the image into two components: the structure and texture images and then repair the two components separately according to their image characteristics. The texture image is repaired by employing the exemplar-based method, for the method is derived from texture synthesis techniques.
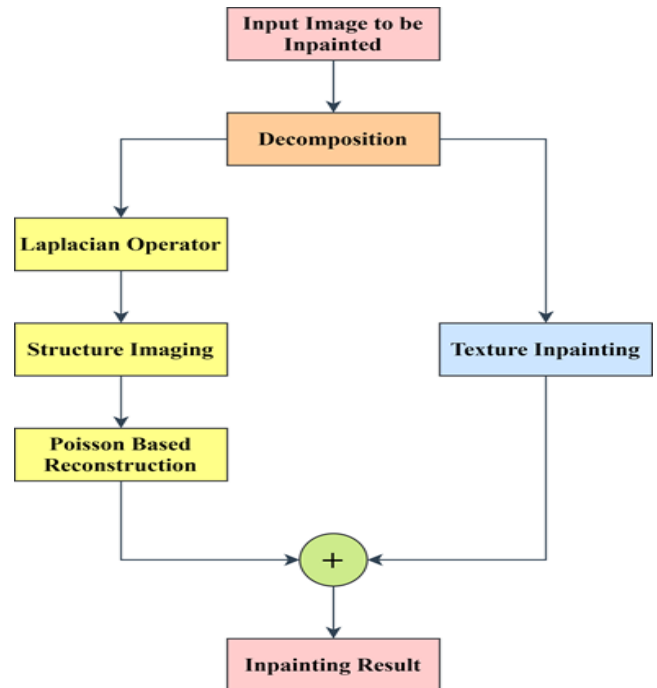

Figure 5 Framework for Inpainting technique.

Figure 5 as for the structure image, first we apply the Laplacian operator to enhance the structure information and obtain the Laplacian image. Then the exemplar-based method is exploited to repair the Laplacian image. Afterward, we reconstruct the structure image from the repaired Laplacian image by the Poisson equation and therefore obtain the repaired result of the structure image. Finally, these two sub-images are added back together to obtain the inpainting result of the original image. By this means, both structure and texture information is well restored.

As to validate our model properly, we have pre-processing steps. We create high quality, the balanced dataset for testing and validation as our downloaded dataset was unbalanced. We used to remove noise in pre-processing that was mainly due to bad light, noise, and air bubbles. These steps (conversion into greyscale, morphological filtering, conversion into a Black hat, applies threshold, Inpaint) provide us clear enhancement to further steps. The pre-processed results are shown in Figure below
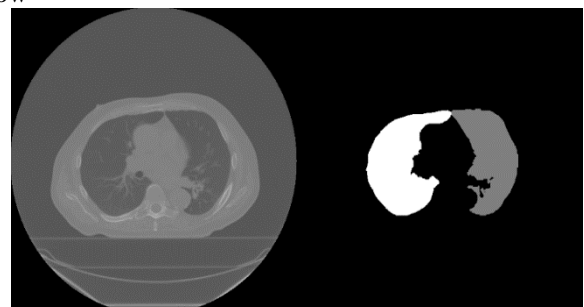

Figure 6 Pre-processed Data.

To improve the learning capability of the model and to generalize its performance, image data augmentation is used. Augmentation is a method that is used to expand artificially the size of a training dataset by creating modified versions of images in the dataset. After the Data I/O Interface initialization, the Data Augmentation class can be configured. A Pre-processor object with default parameters automatically initialize a Data Augmentation class with default values, but here we initialize it by hand to illustrate the exact workflow of the methodology. The parameters for the Data Augmentation configure which augmentation techniques should be applied to the data set. In this case, we are using all possible augmentation techniques to run extensive data augmentation and avoid overfitting.

III. EXPERIMENTAL RESULTS

The proposed automatic detection of kidney diseases in the image is based on fastAI method, which is used to extract the deep features and texture from the image dataset. We want to perform a 3-fold cross-validation for kidney tumor segmentation on the (KITS19) dataset. To achieve this, we are using our medical image segmentation techniques with Convolutional Neural Networks and Deep Learning. The aim of this research is to provide a fast building of medical image segmentation pipelines including data Input/Output, pre-processing, data augmentation, patch-wise analysis, metrics, a library with state-of-the-art deep learning models and model utilization like training, prediction as well as fully automatic evaluation (e.g., cross-validation). Our technique is based on Keras with Tensorflow as backend. The model formulation is used to calculate the parameters and output results.

2D image segmentation, results for each case were analyzed in three dimensions, with specific criteria for inclusion and exclusion in the final volumetric segmentation. Kidney regions with tumors of different types were validated depending upon conditions of volume and position. The ground truth and prediction of our build model is shown in figure below:
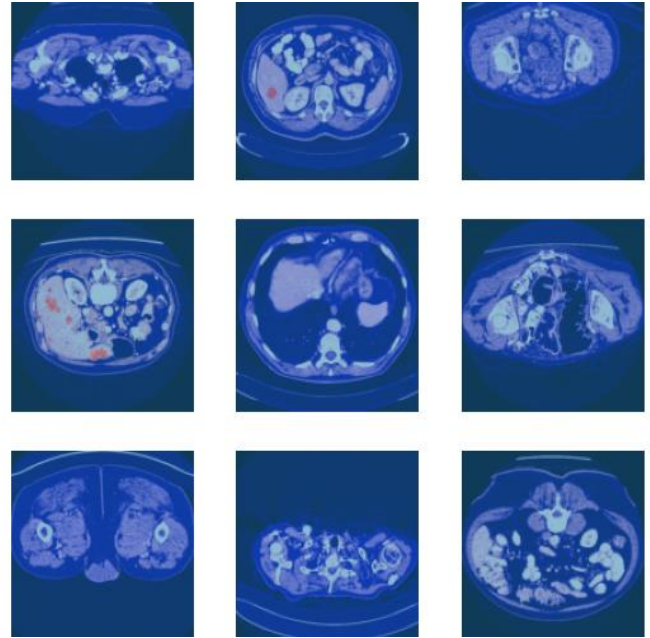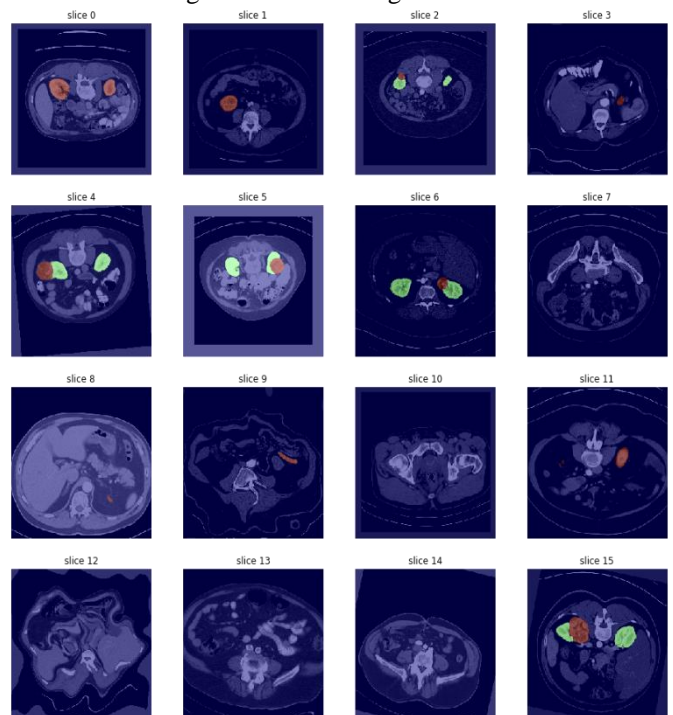

Figure 7 Without Segmentation.

Figure 8 Segmentation and Detection of Tumor after 3 folds of Cross Validation

Prediction of the ground truth of different cases with slice 0 is shown in Figure 9.
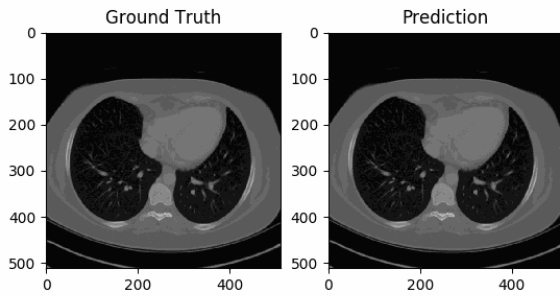
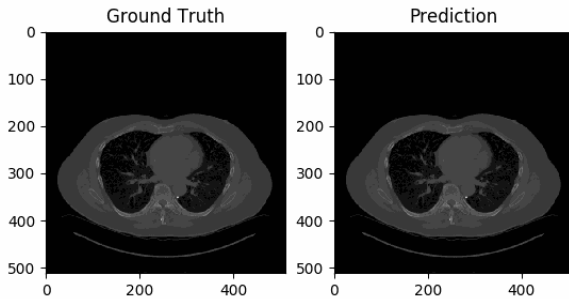Figure 9 Ground Truth and Prediction of Case 24 with Slice 0.

Figure 10 Ground Truth and Prediction of Case 44 with Slice 0.

Let us look at some of the samples that were most difficult for our model to detect. Figure 9 shows some of the random samples that were difficult for our model to detect. At every sub-image and some parameters are shown. The model predicted it as Malignant Tumor Detection, however it was Benign. The losses occur during the prediction of this random image 0.01 and the probability to predict this random sample was 0.

We provided some random images one by one to the model. Since our focus was to detect Kidney tumors. The 'prediction interval' guesses in which value a future individual reflection will fall, whereas a 'confidence interval' displays the likely value of ranges related to the statistical constraint of the information, i.e., the mean of the population. Our model does not only detect Tumor Detection, but it can also detect other Kidney tumors with an impressive less loss. Figures below shows the performance of our model.
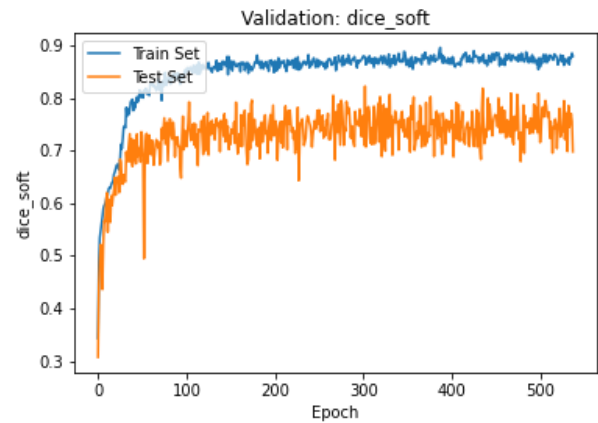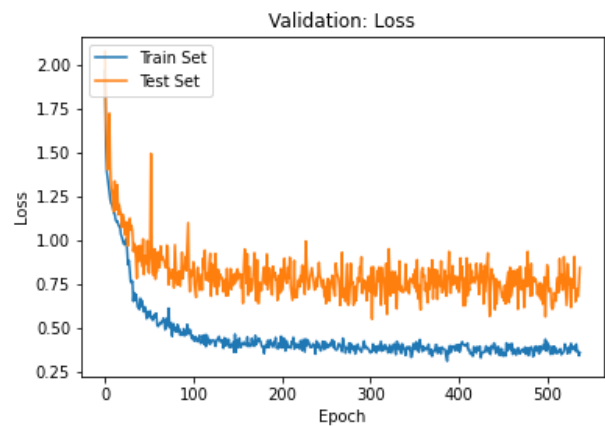

Figure 11 Validation Score of our model CV-1

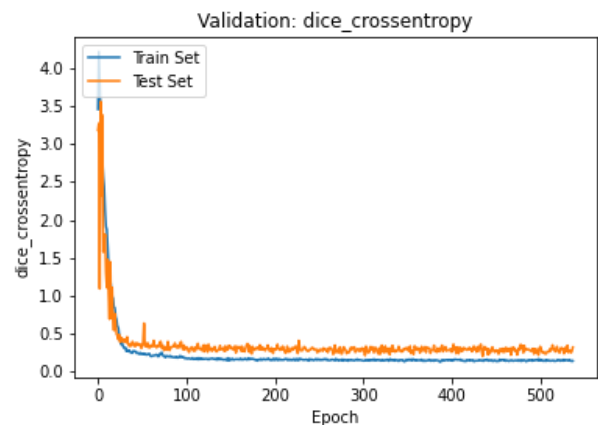Figure 12 Validation Loss of Our Model CV-1

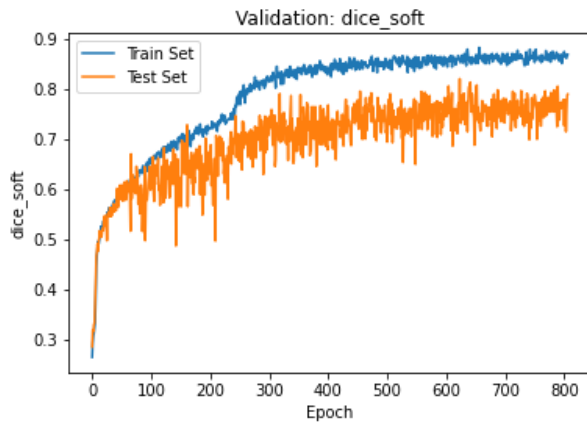Figure 13 Validation Cross-entropy of Our Model CV-1
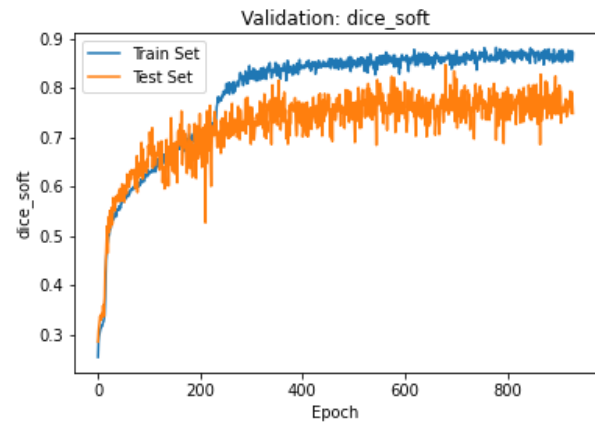
Figure 14 Validation Loss of Our Model CV-2

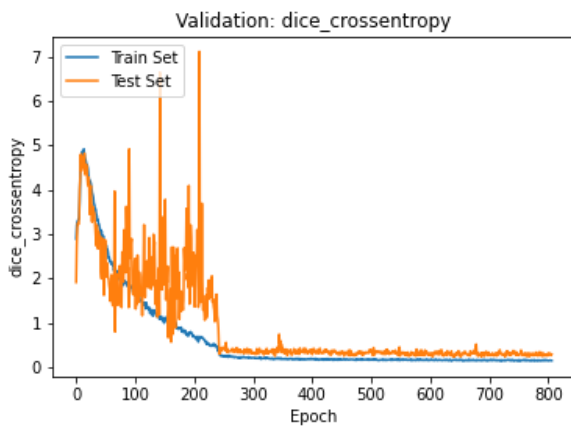Figure 17 Validation Loss of Our Model CV-3

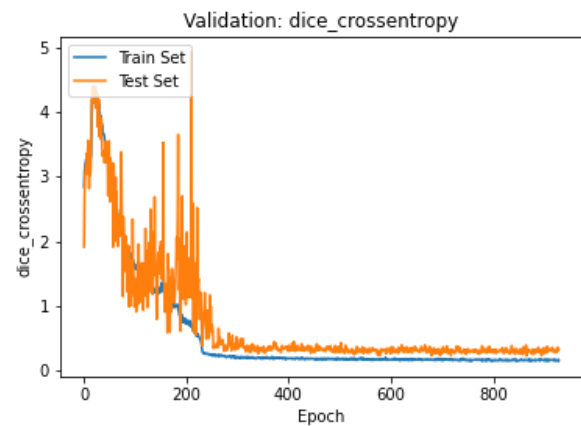Figure 15 Validation Cross-entropy of Our Model CV-2

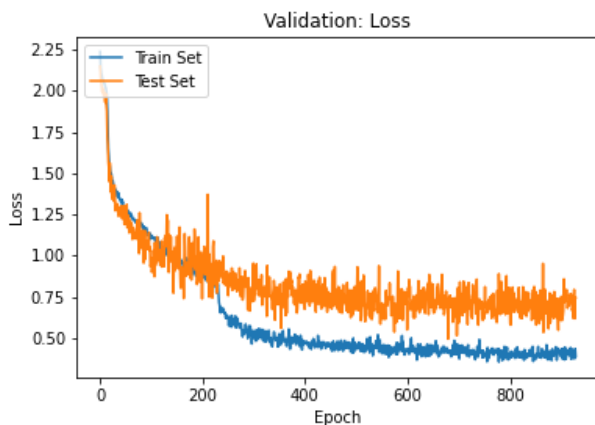Figure 18 Validation Cross-entropy of Our Model CV-3

Figure 16 Validation Score of our model CV-3

In the figures 12 to 18 it can easily be seen that Validation Loss in all the folds is decreasing eventually with respect to time. The quantitative analysis of ROC was used in different publications. The graph shows the comparison of the area under the ROC curve (AUC) and the Equal Error Rate (EER) of different methods. Higher AUC and smaller EER are good. Most papers did not publish AUC/EER

IV.   CONCLUSION

In this research work, we had an image of a CT-Scan of a kidney containing a tumor. We applied the Morphological Filtering for the Image Enhancement followed by Histogram Equalization. The restored image went under Image Segmentation, namely, Watersheds after which Marking was done and the final image was produced. The final image showed a distinct location of the stone in the kidney. Hence the stone was detected. We used the CNN model for testing our sets. We train the model on our training data and found the learning rate of the model. Then we estimate the confusion matrix and accuracy of the trained model. Accuracy of our trained model range from 90-99 percent and are considered as the best accuracies. The model predicted the Tumor Detection with confidence level ranges from 70-100 percent depending on the resolution of the image. The model

also successfully predicted and classify other Kidney tumors with a very high confidence. For future work, one can suggest the implementation of this automated method for the detection of Kidney tumors into smartphones/Tabs. We can also optimize the learning rate of the classifiers. Kidney images can be sliced into layers so that more accurate detection may perform at higher accuracy.

V. REFERENCE

[1] C. Sabanayagam *et al.*, "Articles A deep learning algorithm to detect chronic kidney disease from retinal photographs in community-based populations," *Lancet Digit. Heal.*, vol. 2, no. 6, pp. e295–e302, doi: 10.1016/S2589-7500(20)30063-7.

[2] Parul Sinha and Poonam Sinha, "Comparative Study of Chronic Kidney Disease Prediction using KNN and SVM," *Int. J. Eng. Res.*, vol. V4, no. 12, pp. 608–612, 2015, doi: 10.17577/ijertv4is120622.

[3] J. Norouzi, A. Yadollahpour, S. A. Mirbagheri, M. M. Mazdeh, and S. A. Hosseini, "Predicting Renal Failure Progression in Chronic Kidney Disease Using Integrated Intelligent Fuzzy Expert System," *Comput. Math. Methods Med.*, vol. 2016, 2016, doi: 10.1155/2016/6080814.

[4] K. Kumar and A. Abhishek, "Artificial Neural Networks for Diagnosis of Kidney Stones Disease," *Int. J. Inf. Technol. Comput. Sci.*, vol. 4, no. 7, pp. 20–25, 2012, doi: 10.5815/ijitcs.2012.07.03.

[5] S. Yin *et al.*, "Automatic kidney segmentation in ultrasound images using subsequent boundary distance regression and pixelwise classification networks," *arXiv*, Nov. 2018.

[6] S. Revathy, B. Bharathi, P. Jeyanthi, and M. Ramesh, "Chronic kidney disease prediction using machine learning models," *Int. J. Eng. Adv. Technol.*, vol. 9, no. 1, pp. 6364–6367, 2019, doi: 10.35940/ijeat.A2213.109119.

[7] M. S. Gharibdousti, K. Azimi, S. Hathikal, and D. H. Won, "Prediction of Chronic Kidney Disease using data mining techniques," *67th Annu. Conf. Expo Inst. Ind. Eng. 2017*, no. May 2017, pp. 2135–2140, 2017.

[8] C. Bradshaw *et al.*, "Early detection of chronic kidney disease in low-income and middle- income countries : development and validation of a point-of-care screening strategy for India," pp. 1–9, 2019, doi: 10.1136/bmjgh-2019-001644.

[9] H. A. Wibawa, I. Malik, and N. Bahtiar, "Evaluation of Kernel-Based Extreme Learning Machine Performance for Prediction of Chronic Kidney Disease," in *2018 2nd International Conference on Informatics and Computational Sciences, ICICoS 2018*, Oct. 2019, pp. 33–36, doi: 10.1109/ICICOS.2018.8621762.

[10] E. H. A. Rady and A. S. Anwar, "Prediction of kidney disease stages using data mining algorithms,"

Informatics in Medicine Unlocked, vol. 15. Elsevier Ltd, p. 100178, Jan. 2019, doi: 10.1016/j.imu.2019.100178.