e-ISSN:2582-7219

# INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 7, July 2024



Impact Factor: 7.521

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707186|

# Transforming Kidney Cancer Detection: Advanced Machine Learning for Early Diagnosis and Precision Care

**Gunasekaran K, Varun Gowda P**

Assistant Professor, Department of MCA, AMC Engineering College, Bengaluru, India

Student, Department of MCA, AMC Engineering College, Bengaluru, India

**ABSTRACT:** The kidney is an organ that is key to the survival of a person, and it keeps the body functioning by removing excess water as well as waste from the blood. Although renal disease and kidney cancer are some of the most serious health issues that are now facing on a global scale, cancer of the kidney is the most dangerous one. Kidney cancer, which is classified by the abnormal growth of cells in kidney tissues, requires precision techniques like machine learning. Machine learning models are featuring a new approach to early detection, the strategy starts with data analysis via CT scans and incorporating the MobileNet architecture to set up a neural network that is highly precise in classifying kidney tumor tissues. With validation of the model completed and an accuracy rate of 99.00%, this model proved its tremendous power to diagnose malignant tumors. The CT dataset, consisting of 2,283 tumor class images and 5,077 normal class images, provides a thorough and neat source for learning and estimating. Besides that, utilized blood test samples as an additional approach to detect kidney problems. ANN training confirmed its effectiveness, and achieved 97% validation accu- racy. The dataset of 400 blood tests includes 26 crucial attributes including age, blood pressure, and glucose levels, so that each patient's information can be studied comprehensively to allow an accurate diagnosis. Moreover, the machine learning models involve advanced features with a user-friendly web interface, allowing the health care professionals to easily submit their data and speed up reporting, thereby generating quick and accurate results. Using machine learning to detect early-stage cancer, hoping to cause a shift in the present clinical approach and ultimately save many lives through timely treatments that will improve the health of patients.

**KEYWORDS**: Machine Learning, Accuracy, F1-Score, Kidney, Kidney Cancer, Supervised Machine Learning, Dataset, Feature Selection, Linear Regression, Neural Networks, Performance Metrics, Precision, Recall, Training, Validation, True Positive Rate, Logistic Regression, MobileNet, Prediction

## I. INTRODUCTION

Diagnosis of kidney cancer is an essential activity in the healthcare infrastructure and usually is done based on image analysis and clinical data. Accuracy concerns and computa- tional efficiency issues have rendered several other systems incompatible. Traditional systems with high processing time and complexity render the decision useless and may cause diagnosis delays or adverse effects on patient diagnosis. To improve on this, the work presents a novel method utilizing advanced machine learning systems. The method focuses on a novel use of MobileNet, an artificial neural network model. In summary, the proposed approach was evaluated us- ing comprehensive datasets of medical images and clinical records. Specifically, the medical imaging dataset consisted of a significant number of CT scan images for normal and tumor cases. Furthermore, the clinical data records contain the patient's blood test results and other important data to allow for a complete understanding of the patient's condition. This approach preprocesses both the medical images and clinical data. For medical images, resizing and normalization are done to ensure They can serve as machine learning input data models. An efficient and effective architecture for image classification, MobileNet, is used to extract the features from the resized images.To optimize face diagnostic accuracy,the ANN model was applied to clinical data in between applications

## II. LITERATURE REVIEW

Analogous research has been conducted on the use of deep learning and computer-assisted to develop an automated kidney cancer detection system.
Kwang-Hyun Uhm had introduced a Convolutional Neural Network (CNN) primarily based model for the

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707186|

comprehensiveanalysis of kidney tumors.The model makes use of 3-d CNN- based segmentation for unique protection; it became trained and assessed on CT records from 308 nephrectomy patients. The model shows incredible accuracy, with an Area Underthe Curve (AUC) of 0.889 for 308 sufferers and 0.855 for 184 sufferers [1].

Ghalib et al had focused on enhancing prediction accuracy, reducing processing time, and minimizing expenses compared to present approaches with 196 CT scan images. The re- search makes use of Artificial Neural Networks (ANN) and preprocessing methods to categorise tumors, and it performs thoroughly, with an accuracy fee of 93.39% and a median execution time of 0.85 seconds [2].

Luana Batista da Cruz have used a unique batch-basedsynthesis set of rules for kidney detection in CT snap shots with 210 CT scan images. With a 9.6% detection charge,the approach suggests high effectiveness in numerical tests. Key overall performance indicators (Kidney detection in CT pics) which includes Dice Coefficient (96.33%), Jaccard Index(93.02%), Sensitivity (97.42%), and Specificity (99.94%) [3].Seokmin Han has used a modern photo-based, totally deep learning method to distinguish clean cellular, papillary, and chromophobe subtypes of renal mobile carcinoma (RCC) in CT scans. The dataset includes 169 patients with kidney cancer kidney cancer in the advanced stage. It is incorporated into the IoMT and utilized in machine learning to address improved diagnostic intellect and prediction capacities for the study of kidney diseases [11].

Nagarajan S. and Ramprasath M. presented an innovative, fully automated, two-stage method consisting of CNN-based segmentation and level-set algorithms for accurate segmenta- tion of the kidney on DCE-MRI scans. The method is tested and compared with the ground truth on real DCE-MRI scansof 45 subjects and has better performance with DSC of 0.95 ± 0.02, defined as intersection over union of 0.91 ± 0.03, keepingthe 5% Hausdorff distance at 1.54 ± 1.6 [12].

Les T paper proposes a new lesion detection approach according morphological cascade Convolution Neural Network (CNN) with several intersection-over-union (IOU) thresholds. The experimental outcomes represent an AP between 0 and 0.84 and an area under curve of 0.871, suggesting that the proposed method performs well in automating the process[13].

Zhao Z presented an approach to accurate segmentation of kidneys and kidney tumors based on a BA-Net with CT intheir study. Multi-scale inputs and deep supervision are usedto effectively trace hierarchical structural details while dealingwith various tumor sizes. In experiments on KiTS19 dataset, BA-Net shows its effectiveness in accurate kidney and kidney tumor segmentation [14].

Nayak S proposed a new approach to segmenting kidney MRI that targets the enhancement of lesion detection using a hard-clustering technique with the Slime Mould Algorithm. Validation experiments in both quantitative and qualitative respects also confirmed the effective performance of our approach in improving lesion detection capability in accurate kidney MRI segmentation [15].

Thomas N.R research concentrates on the application of CAD algorithms in the early diagnosis of renal tumors from CT images. In this process, segmentation is done using the methods of fuzzy C-means clustering and watershed in order toaccurately segment the kidney regions and then remove tumorareas from the segmented images. The next step is to enhance the accuracy of diagnosis in renal tumors and plan proper treatment [16].

The majority of current study focuses on different kidney cancer diseases. There is a shortage in the literature of binary classification techniques for detecting different types of kidneycancer because many of these studies focus on multi-class classification.

## III. METHODOLOGY

*A. Dataset Collection:*

In general, any website able can be applied to gather data. The dataset is collected from kaggle website, This has important details for kidney cancer detection. The dataset comprises 401 individual entries with 26 columns, each representing different attributes such as age, bp (blood pressure), sg (specific gravity), al (albumin), su (sugar), rbc (red blood cells), pc (pus cell), pcc (pus cell clumps), ba (bacteria), bgr (blood glucose random), bu (blood urea), sc (serum creati- nine), sod (sodium), pot (potassium), hemo (hemoglobin),

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707186|

pcv (packed cell volume), wc (white blood cell count), rc (red blood cell count), htn (hypertension), dm (diabetes mellitus), cad (coronary artery disease), appet (appetite), pe (pedal edema), ane (anemia), id which are the various blood param- eters for blood test analysis using artificial neural network. Similarly the dataset is supplemented with kidney CT scan images obtained from tumor and normal patients, totaling 2,283 tumor class images and 5,077 normal images for Ctscan image analysis using mobilenet model showed in Fig 1.

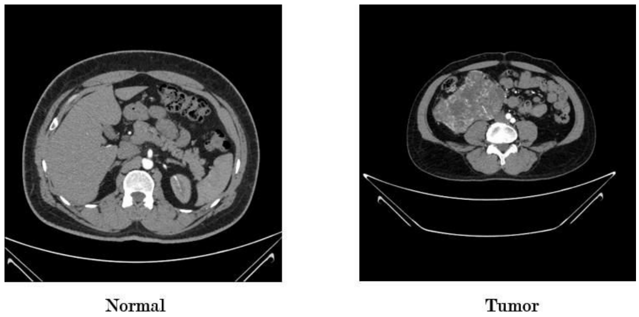


Fig. 1. Images of dataset

*B. Data Preprocessing*

This research exercise paid the necessary attention to the preprocessing steps of the dataset to improve its quality and eligibility for further analysis and modeling. Moreover, during the further data exploration, it was discovered that there Several values were absent in several columns. The categorical values in the column 'class' meaning different conditions of the patients' kidneys, was rationalized to be binary to facilitate subsequent steps in the analysis and model- ing. Additionally, all numerical data columns were thoroughly converted to numeric types to enhance consistency. In thesame line, absent values in the various columns were also well handled. In this case, missing entries in specific columns were filled using the mean values the remainder of the non- missing ones to maintain the integrity and consistency of the data. An important part of the preprocessing is a pro-cess of the conversion of categorical variables to numerical equivalents. The execution of this change is impeccable, with categorical variables such as 'red blood cells', 'pus cell', and 'hypertension' being changed to binary numerical values and become useful for any modeling activity. As well, the final data was put via a procedure of refinements to get it ready for predictive modeling. Unneeded columns, like the identifierand variables that are not to be employed in the selection stage, were eliminated one by one. In the end, this led to the productionof the feature set ('X') and target variable ('y') wanted for training and examining the predictive models.

Similarly after collecting all the pictures and labels, theimages are reorganized to a standard size of (200, 200). The convergence and execution of the model are improved by

TABLE I

_ _ MODEL_ HYPERPARAMETERS

| Model | Training Time (mins) | Trainable Parame- ters | Optimizer | Learning Rate |
|---|---|---|---|---|
| ANN | 2.8 | 2119 | Adam | 0.001 |
| MobileNet | 6.7 | 2050 | Adam | 0.001 |

1. RESULTS

The methods of evaluating results are based on the deter- mination of the level of performance of the working model on information that is hidden or a test dataset. This evaluationis done to determine whether the model can be appliedto new and unknown instances and recognize any possible technical limits of the model such as overfitting or underfitting. Several metrics are accustomed to compare the models' respective performances, including accuracy, Precision, Recall,and F1- Score.

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707186|

*A. Accuracy*

Accuracy of the machine learning models is the most important metric used to assess the effectiveness of models which appears in Table II. It shows the number of times whenthe models perform well versus the whole number of inputs. The application of the two models, i.e., the Artificial Neural Network (ANN) model, and the MobileNet model, has generated two sets of accuracy predictions, for the training and validation datasets separately. The validation the ANN model's accuracy was seen to be 97.5%, which is akin totelling that the model yields a precision of 97.5% on thedata that it had no contact with during the training. Furtheron, for the MobileNet model the training accuracy as well as validation accuracy were also evaluated. From validationit was noteworthy that precision was far greater at 99.52%. Accuracy values help to assess how much models do their jobsby the data which they received training for upon additionally data they did not see before. Therefore, these accuracy levels are needed when the models are used for practical applications.

TABLE II ACCURACY

| Model | Validation Accuracy |
|---|---|
| ANN | 97.5% |
| MobileNet | 99% |

*Precision, Recall, and F1-Score*

Further, precision is an additional KPI that is utilized to assess the operation of machine learning systemsfor tasks such as classification which is shown in table III.It measures the precision that is connected with the cor- rectness of positive predictions of the model, which means the percentage of all forecasts that come true that are the correct ones out of the total number of positive predictions made. With the setting up of two models in the subsequent context,precision was figured out for each class individually. For the Artificial Neural Network (ANN) model, precision values were determined as follows: precision for class 0 (negative class) was 0.93, while the precision for class 1 (positiveclass) was
1.00. This precision measure indicates model's capacity to correctly label for each class and hence evaluate the same. More precisely, model demonstrates that the accuracy rate of the class0 was 0.93, which hints about 93% of instances predicted as class 0 were really negatives. Likewise, a precision of 1.00was identified for class 1, showing that the entire disruptive detection cast as class 1 were, actually true positives.
When it comes to the MobileNet model, Additionally, accuracy was calculated for each class: class 0 accuracy was 95%, while class 1 (positive class) precision was 98%. These values of precision indicate that the MobileNet model's recognition capability is strong enough for the perfect categorization ofthe elements of each class. The accuracy of the model is95% for class 0, and class 1 has precision of 98%. This shows high degree of accuracy in positive classification of bothclasses by using the MobileNet model. This provides essentialinformation to also complement the accuracy and recall.

TABLE IIIPRECISION

| Model | Class(Negative) | Class (Positive) |
|---|---|---|
| ANN | 0.93 | 1.0 |
| MobileNet | 0.95 | 0.98 |

TABLE IVRECALL

| Model | Class(Negative) | Class (Positive) |
|---|---|---|
| ANN | 1.0 | 0.96 |
| MobileNet | 0.98 | 0.99 |

The recall values for both models from the table IV capture their success in the classification subproblem, where they correctly classify stories as being either positive or negative. For the negative class, The ANN model has a noticeably superior performance and resulted in the recall being 1.00, while for thepositive class the recall was 0.96. It follows

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707186|

that, the ANN actually captured the complete set of the negative class, with 96% of instances being correctly categorized as belonging to the positive class. Likewise, the recall scores for the MobileNet model – including the principles of 0.98 for negative class and 0.99 for positive class – came out to be notablyhigh. Evaluation findings derived from this metric indicate that our MobileNet model displays high confidence in correctly identifying both classes, particularly positive ones.

The F1-score ,which is the harmonic mean of precision and recall, which considers both false positives and false negatives,provides a holistic evaluation by capturing both the false positives and false negatives aspects of prototype is seen in table V. For ANN, the model obtained an F1-score of 0.97 for the unfavorable category and 0.98 for the positive class. This quiver demonstrate a solid weight of between accuracy and memory and consequently demonstrates that the model is precisely maximize on both the false negatives and the positives. However, the MobileNet model revealed some degree of superiority regarding F1-scores, namely, 0.98 negative class score and a 0.99 positive class score. The above outcomes signify, that is to say, that MobileNet is great at precision as well as recall across different classes, depicting its fault-tolerant capability of correctly classifying instances with minimum error.

TABLE VF1-SCORE

| **Model** | **Class(Negative)** | **Class (Positive)** |
|---|---|---|
| ANN | 0.97 | 0.98 |
| MobileNet | 0.98 | 0.99 |

*B. AUROC Curve*

The Fig 2 represents the ROC (Receiver Operating Char- acteristics) curve, which plots the true positive rate (TPR) versus the false-positive rate (FPR) for each class considering two classes with the area under the ROC curve (AUC) scores attached to those curves. Similarly in Fig 3 both ROC and AUROC curve for class 0 as well are calculated. The optimal ROC curves are shown and the value of AUROC score is presented for class 0. Such mixed units allow the user totrack the overall activity of the classifier and from “false” ones. This policy provides condition under which the model

performance capability could be approximated in binaryclassification tasks.distinguish “true”positive

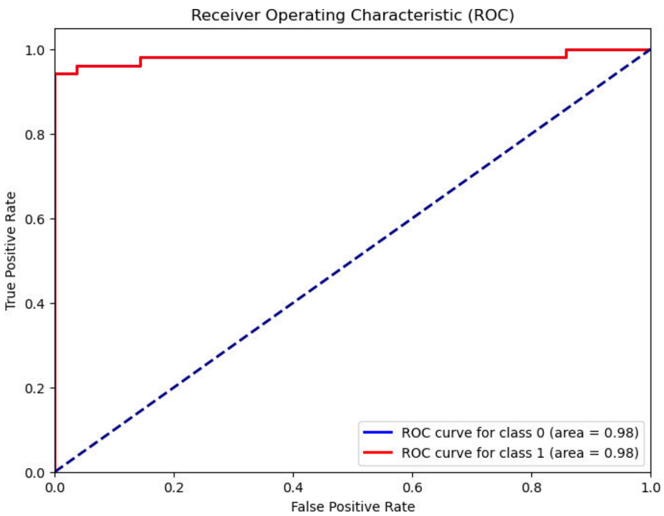


Fig. 2. AUROC curve of ANN model

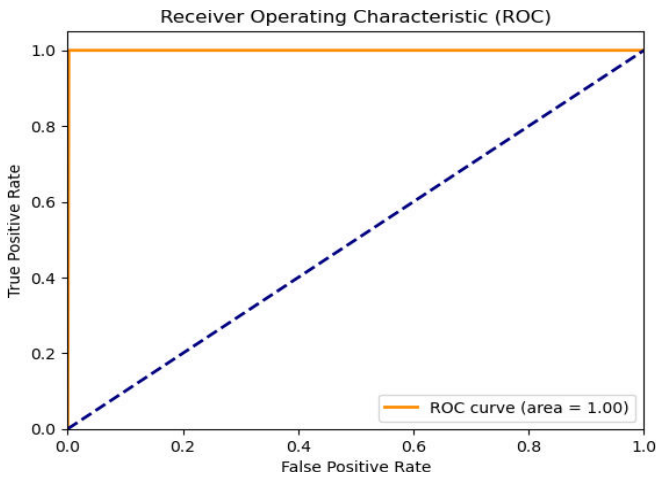


Fig. 3. AUROC curve of MobileNet model

## IV. CONCLUSION

Comprehensive kidney cancer detection system represents a significant advancement in healthcare technology that makes use of deep learning models and an easy-to-use online inter- face and provides a comprehensive detection

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707186|

approach. The ANN and the MobileNet models did, in fact, give us satis- factory results when they acted as cancer and detectors. Forthe ANN, the runtime was 97.5%, while MobileNet's accuracyscore was 99.5%. Furthermore, both methods showed a high degree of precision, recall and F1-score values, reflecting the efficiency of the models into diagnosing kidney CT scan images. For the next stage, the researchers might develop the models so that they could cover a collection of images with different themes, proof and fine-tune the thresholds for the ultimate model performance, and use the ensemble techniques to take advantage of strong points in the models.

## REFERENCES

1. Uhm, K.H., Jung, S.W., Choi, M.H., Shin, H.K., Yoo, J.I., Oh, S.W., Kim, J.Y., Kim, H.G., Lee, Y.J., Youn, S.Y. and Hong, S.H:"Deep learning for end-to-end kidney cancer diagnosis on multi-phase abdominal computed tomography" NPJ precision oncology, 2021.
2. Gharaibeh, M., Alzu'bi, D., Abdullah, M., Hmeidi, I., Al Nasar, M.R., Abualigah, L. and Gandomi, A.H: "Radiology imaging scans for early diagnosis of kidney tumors: a review of data analytics-based machine learning and deep learning approaches" Big Data and Cognitive Computing, 2022.
3. da Cruz, L.B., Arau´jo, J.D.L., Ferreira, J.L., Diniz, J.O.B., Silva, A.C., de Almeida, J.D.S., de Paiva, A.C. and Gattass, M: "Kidney segmentation from computed tomography images using deep neural network. Computers in Biology and Medicine" 2020.
4. Han, S., Hwang, S.I. and Lee, H.J: "The classificationof renal cancer in 3-phase CT images using a deep learning method" Journal of digital imaging 2019.
5. Kang, L., Zhou, Z., Huang, J., Han, W. and Member, I.E.E.E: "Renal tumors segmentation in abdomen CT Images using 3D-CNN and ConvLSTM" Biomedical Signal Process- ing and Control, 2022.
6. Akram Z, Kareem MS, Mughal B, Ahmed Z, and Aziz
7. S. "Cancerous Tumor Segmentation of Kidney Images and Prediction of Tumor Using Medical Image Segmentation and Deep Learning Techniques", Published: 09 Mar 2021.
8. Ehwa Yang, Chan Kyo Kim, Yi Guan, Bang-Bon Koo, Jae-Hun Kim, "3D multi-scale residual fully convolutional neural network for segmentation of extremely large-sized 25 kidney tumor", 3 January 2022.
9. Seokmin Han, Sung Il Hwang, Hak Jong Lee, "The classification of renal cancer in 3- phase CT images using a deep learning method", Published: 16 May 2019.
10. Fuzhe Ma a, Tao Sun a, Lingyun Liu b, Hongyu Jing c, "Detection and diagnosis of chronic kidney disease using deep learning-based heterogeneous modified artificial neural network", Published: 24 April 2020.
11. Andriy Myronenko, Ali Hatamizadeh, "3d kidneys andkidney tumor semantic segmentation using boundary-aware networks", 14th September 2019.
12. Chen G., Ding, C., Li, Y., Hu, X., Li, X., Ren, L., Ding, X., Tian, P. and Xue, W., 2020. Prediction of chronic kidney disease using adaptive hybridized deep convolutional neural network on the internet of medical things platform.
13. Nagarajan S. and Ramprasath M., 2023, November. Enhanced study on Deep learning model for kidney segmen- tation using DCE-MRI. In 2023 International Conference on Research Methodologies in Knowledge Management, Artifi- cial Intelligence and Telecommunication Engineering (RMK- MATE) (pp. 1-5). IEEE.
14. Les, T., Markiewicz, T., Dziekiewicz, M. and Lorent, M., 2018, September. Automatic recognition of the kidney in CT images. In 19th International Conference Computational Problems of Electrical Engineering (pp. 1-4). IEEE.
15. Zhao, Z., Chen, H., Li, J. and Wang, L., 2022, July. Boundary attention u-net for kidney and kidney tumor segmentation. In 2022 44th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 1540-1543). IEEE.
16. Si, T., Nayak, S. and Sarkar, A., 2021, December. Kidney MRI segmentation for lesion detection using cluster- ing with slime mould algorithm. In 2021 10th international conference on internet of everything, microwave engineering, communication and networks (IEMECON) (pp. 01-06). IEEE.
17. Thomas, N.R. and Anitha, J., 2022, June. An automated kidney tumour detection technique from computer tomography images. In 2022 International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS)(pp. 1-6). IEEE.