# Enhanced Transfer Learning Strategies for Effective Kidney Tumor Classification with CT Imaging

Muneer Majid[1]*, Yonis Gulzar[2]*, Shahnawaz Ayoub[3], Farhana Khan[4],
Faheem Ahmad Reegu[5], Mohammad Shuaib Mir[6], Wassim Jaziri[7], Arjumand Bano Soomro[8]
Glocal School of Science and Technology, Glocal University, Delhi-Yamunotri Marg (State Highway 57),
Mirzapur Pole, Dist - Saharanpur, U.P. - 247121, India[1, 3, 4]
Department of Management Information Systems, College of Business Administration,
King Faisal University, Al-Ahsa 31982, Saudi Arabia[2, 6, 7, 8]
College of Computer Science and Information Technology, Jazan University, Jazan 45142, Saudi Arabia[5]
Department of Software Engineering-Faculty of Engineering and Technology, University of Sindh, Sindh, Pakistan[8]

***Abstract*—Kidney tumours (KTs) rank seventh in global tumour prevalence among both males and females, posing a significant health challenge worldwide. Early detection of KT plays a crucial role in reducing mortality rates, mitigating side effects, and effectively treating the tumor. In this context, computer-assisted diagnosis (CAD) offers promising benefits, such as improved test accuracy, cost reduction, and timesaving compared to manual detection, which is known to be laborious and time-consuming. This research investigates the feasibility of employing machine learning (ML) and Fine-tuned Transfer Learning (TL) to improve KT detection. CT images of individuals with and without kidney tumors were utilized to train the models. The study explores three different image dimensions: 32x32, 64x64, and 128x128 pixels, employing the Grey Level Co-occurrence Matrix (GLCM) for feature engineering. The GLCM uses pixel pairs' distance (d) and angle (θ) to calculate their occurrence in the image. Various detection approaches, including Random Forest (RF), Support Vector Machine (SVM), Gradient Boosting (GB), and Light Gradient Boosting Model (LGBM), were applied to identify KTs in CT images for diagnostic purposes. Additionally, the study experimented with fine-tuned ResNet-101 and DenseNet-121 models for more effective computer-assisted diagnosis of KT. Evaluation of the efficient diagnostics of fine-tuned ResNet-101 and DenseNet-121 was conducted by comparing their performance with four ML models (RF, SVM, LGBM, and GB). Notably, ResNet-101 and DenseNet-121 achieved the highest accuracy of 94.09%, precision of 95.10%, recall of 93.5%, and F1-score of 93.95% when using 32x32 input images. These results outperformed other models and even surpassed state-of-the-art methods. This research demonstrates the potential of accurately and efficiently classifying KT in CT kidney scans using ML approaches. The use of fine-tuned ResNet-101 and DenseNet-121 shows promising results and opens up avenues for enhanced computer-assisted diagnosis of kidney tumors.***

***Keywords*—Kidney; kidney tumor; automatic diagnosis; machine learning algorithms; CT imaging; deep learning; transfer learning***

## I. Introduction

The kidneys, two bean-shaped organs located on either side of the spine, are indispensable for maintaining the body's internal equilibrium and overall health. These vital organs play a pivotal role in filtering waste products, excess salts, and water from the blood, while also regulating blood pressure and producing essential hormones for red blood cell production. With such critical functions, any disruption or abnormality in kidney health can have serious repercussions on an individual's well-being [1]. Kidney tumors can be benign or malignant [2]. The benign kidney tumor can be a cyst, masses, or lipoma whereas malignant kidney tumor refers to renel cancer, pelvic cancer [3]. Kidney cancer, characterized by the presence of tumors arising from renal tissues, poses a significant global health challenge. Despite extensive research, the precise etiological factors triggering kidney cancer remain enigmatic, though hereditary and environmental influences are among the potential contributors. The insidious nature of kidney tumors often leads to asymptomatic progression, delaying diagnosis until advanced stages.

In 2023, it is estimated that 81,800 adults in the United States will be diagnosed with kidney cancer. Kidney cancer is more common in men, and the average age of diagnosis is 64, with most cases occurring between ages 65 and 74. The number of new kidney cancer cases has been increasing, though the rate of increase has slowed in recent years, partially due to increased use of imaging tests that can detect small kidney tumors incidentally [4]. In 2020, an estimated 179,368 people worldwide died from kidney cancer. The 5-year relative survival rate for kidney cancer in the United States is 77%. This rate varies depending on cancer stage, age, general health, and treatment effectiveness. For instance, the five-year relative survival rate is 93% for those with cancer confined to the kidney, 72% if cancer has spread to surrounding tissues or lymph nodes, and 15% if it has spread to distant parts of the body[4].

In the realm of medical imaging, Computed Tomography (CT) scans, ultrasounds, and Magnetic Resonance Imaging (MRI) serve as vital techniques, affording physicians a comprehensive visualization of the kidneys and associated tumors, thereby enabling meticulous evaluation of their dimensions, morphology, and characteristics [5,6]. Despite their undeniable utility, distinguishing between healthy tissue and malignant growth in kidney scans poses a formidable challenge. While manual diagnosis and expert detection of kidney tumors boast commendable accuracy, they demand significant time and effort [7], and results may exhibit

*Corresponding Author: ygulzar@kfu.edu.sa, muneerbazaz@gmail.com

variability amongst different practitioners. The paramount significance of early detection and accurate classification of tumors lies in mitigating the risk of metastasis to other anatomical regions. As an expedient and efficient alternative, automatic detection emerges as a promising approach, streamlining diagnostic processes and potentially safeguarding patients' lives. Although manual detection is renowned for its precision, the automatic diagnostic methodology offers expedited results, without compromising comparability to manual findings.

Artificial intelligence (AI) and deep learning have revolutionized various industries, including agriculture [8–13], education [14, 15], finance [16], and healthcare [17–19]. In the field of healthcare, AI has shown tremendous promise in improving patient outcomes, enhancing diagnostics, and streamlining healthcare processes. With the ability to analyze vast amounts of data and identify complex patterns, AI-powered systems have opened new frontiers for early disease detection, personalized treatment plans, and overall healthcare efficiency. In healthcare, one of the areas where AI and deep learning have made significant advancements is in the early detection of diseases, including cancer [20]. Detecting cancer at an early stage is crucial for improving treatment success and patient survival rates. Kidney cancer, for example, often presents with few symptoms in its early stages, making early detection challenging. However, deep learning algorithms have proven to be effective in analyzing medical imaging data, such as CT scans and MRI images, to detect kidney tumors at their nascent stages [21].

From the literature, it is evident that many researchers have incorporated deep learning in classifying kidney tumors. Lee et al. [22] developed an automatic deep feature classification (DFC) method using hand-crafted and deep features, along with machine learning classifiers, to distinguish benign angiomyolipoma without visible fat (AMLwvf) from malignant clear cell renal cell carcinoma (ccRCC) in abdominal contrast-enhanced computer tomography (CE CT) images. The proposed method achieved an accuracy of 76.6 ± 1.4% using the combination of hand-crafted and deep features, outperforming HCF-only and DF-only methods by 6.6%p and 8.3%p, respectively. Texture image patches (TIPs) were introduced to emphasize texture information and reduce mass size variability, resulting in steady performance regardless of the convolutional neural network (CNN) models used. Han et al. [23] used an image-based deep learning framework to classify renal cell carcinoma subtypes using CT images. The neural network achieved 0.85 accuracy, 0.64-0.98 sensitivity, 0.83-0.93 specificity, and 0.9 AUC, showing promising results for subtype classification and potential clinical cooperation with radiologists. Deep learning framework achieved 93.39% accuracy in classifying clear cell RCC and 87.34% for chromophobe RCC from histopathological images. A novel support vector machine-based method improved classification accuracy to 94.07% for distinguishing clear cell, chromophobe, and papillary RCC. The CNN also extracted morphological features to predict patient survival outcome, showing potential for cancer diagnosis and prognosis [24]. Oberai at al. [25] developed a semi-automated majority voting CNN-based method to classify renal cell carcinoma (RCC) from benign solid renal masses on contrast-enhanced computed tomography (CECT) images. The CNN model achieved 83.75% accuracy in differentiating RCC from benign masses. A fully automated approach yielded 77.36% accuracy, while a 3D CNN achieved 79.24% accuracy in renal mass classification. Pedersen et al. [26] used a modified version of ResNet50V2 CNN to differentiate oncocytoma from renal cell carcinoma (RCC) using non-invasive imaging. They collected 20,000 2D CT images from 369 patients for training, validation, and testing. The model achieved 93.3% accuracy and 93.5% specificity on the main test set and 90.0% accuracy and 98.0% specificity on the additional validation set. When evaluated with a majority vote for each patient, the accuracy rose to 100%, reducing false negatives to zero, demonstrating the potential of CNNs for accurate diagnosis.

Sudharson et al. [27] proposes an automatic classification method for B-mode kidney ultrasound images using an ensemble of deep neural networks (DNNs) with transfer learning. The DNNs, including ResNet-101, ShuffleNet, and MobileNet-v2, are combined using majority voting for better classification performance. The method achieves a maximum classification accuracy of 96.54% for quality images and 95.58% for noisy images, outperforming existing methods, making it a valuable tool for precise diagnosis of kidney diseases. Pirmoradi et al. [28] proposes a new machine learning approach for identifying significant miRNAs and classifying kidney cancer subtypes to develop an automatic diagnostic tool. The method involves two main steps: feature selection using the AMGM measure to choose candidate miRNAs and classification using a self-organizing deep neuro-fuzzy system, which overcomes challenges in high-dimensional data analysis. The results shows that the proposed method achieves high accuracy in classifying kidney cancer subtypes based on the selected miRNAs. Abdeltawab et al. [29] proposes a deep learning pipeline for automated classification of kidney cancer subtypes, specifically clear cell renal cell carcinoma and clear cell papillary renal cell carcinoma. The model uses convolutional neural networks on whole slide images divided into patches, providing patchwise and pixelwise classification. The approach accurately classifies the four classes and outperforms other state-of-the-art methods. This deep learning method has the potential to assist pathologists in diagnosing kidney cancer subtypes from histopathological images. Abdeltawab et al. [30] presents a deep learning framework for classifying kidney tumor subtypes, specifically clear cell renal cell carcinoma and clear cell papillary renal cell carcinoma. The framework utilizes three convolutional neural networks to process kidney image patches of different sizes, providing patchwise and pixelwise classification. The results demonstrate superior performance compared to existing methods, highlighting the potential of deep learning techniques in cancer diagnosis. Khan et al.[31] addressed urgent brain tumor diagnosis using automated methods, specifically convolutional neural networks (CNNs). Deep convolutional features improve classification accuracy significantly, and an ensemble of XGBoost, AdaBoost, and Random Forest achieves a top accuracy of 95.9% for tumors and 94.9% for normal cases, surpassing individual methods. Zhu et al. [32] presents a pipeline employing transfer learning

to address the limitations of small medical image datasets in deep learning applications. The proposed dual-channel fine segmentation network (FS-Net) effectively segmented kidney and tumor regions in 3D CT images, outperforming state-of-the-art methods. The classification model using radiomics features demonstrated accurate classification of benign and malignant tumors in the small dataset. The work emphasizes the significance of architecture design in transfer learning and provides valuable insights for small data analysis in medical imaging. Gulzar at al. [33] conducted using CT scans of 125 subjects, reveals a negative correlation between visceral fat to abdomen size ratio and mean liver intensity values, as well as between mean liver intensity values and total abdomen fat to abdomen size ratio. These correlations indicate a direct link between obesity and diffuse liver fat. This insight contributes to understanding fatty liver disease and its associated health risks. Sarada et al. [34] proposes a hybrid ensemble of visual capsule networks and deep feed-forward extreme learning machines for kidney tumor classification and segmentation from CT images. The model achieves a DICE coefficient of 0.96 and an accuracy of 97.5%, outperforming other hybrid deep learning models. Zhao et al. [35] purpose of the study was to develop an automated method using 3D U-Net and ResNet for accurate segmentation and classification of renal masses in CT images. The algorithm achieved high performance in kidney boundary segmentation (Dice coefficient of 0.99) and renal mass delineation (average Dice coefficients of 0.75 and 0.83). The classification accuracy for masses was 86.05% for masses <5 mm and 91.97% for masses ≥5 mm. The proposed method demonstrated the capability of accurately localizing and classifying renal masses.

In this study, proposes a computer-assisted diagnosis system using machine learning and fine-tuned transfer learning for efficient kidney tumor detection in CT images, achieving high accuracy using different deep learning models. The contribution of this study is as follows:

- Enhanced Kidney Tumor Detection: The study introduces a computer-assisted diagnosis (CAD) system utilizing machine learning and fine-tuned transfer learning, which improves the accuracy and efficiency of kidney tumor detection in CT images compared to manual methods.
- ML Model Evaluation: The research comprehensively experiments with various machine learning models, including Random Forest, Support Vector Machine, Gradient Boosting, and Light Gradient Boosting Model, to identify kidney tumors in CT images. The comparison of these models helps identify the most effective approach for tumor detection.
- Fine-tuned Transfer Learning: The study proposes and evaluates the performance of fine-tuned ResNet-101 and DenseNet-121 models for computer-assisted diagnosis of kidney tumors. These models demonstrate superior accuracy, precision, recall, and F1-score compared to other models and state-of-the-art methods.

Efficient Data Preprocessing: The research applies different pre-processing techniques and image resizing to reduce the complexity of the training model and speed up the training process. This optimization ensures faster and more efficient kidney tumor classification.

## II. Material and Methods

This section provides an elaborate exposition of the proposed framework, incorporating established machine learning (ML) algorithms and deep learning models using transfer learning (TL). As depicted in Fig. 1, the framework encompasses two core components: data pre-processing and the scaling of input image dimensions. Subsequently, a diverse set of ML and TL models were trained on the dataset, comprising Chest Computed Tomography (CT) images.

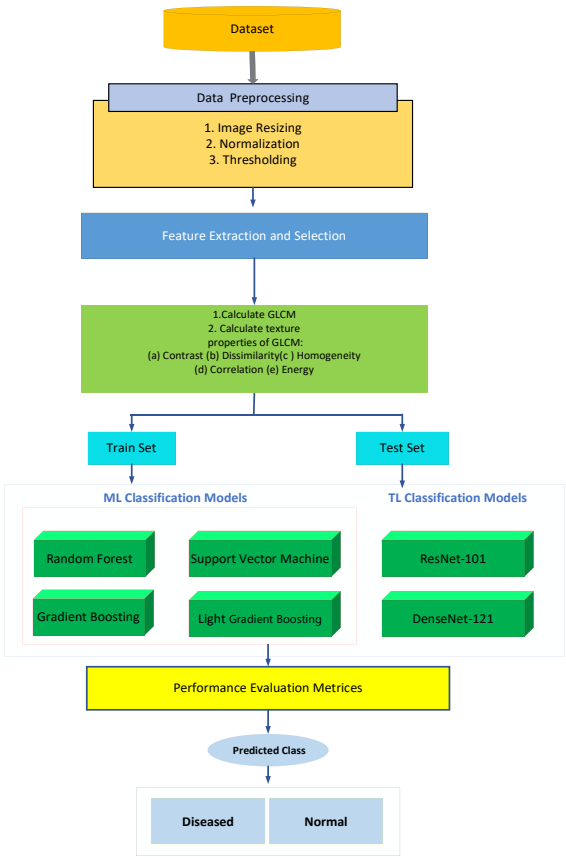


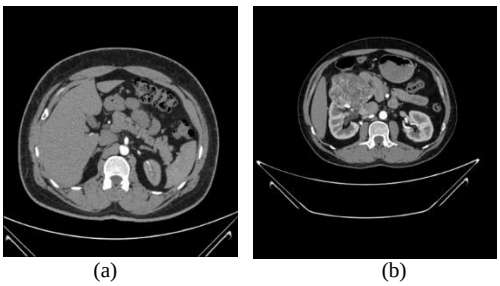

Fig. 1. Detailed framework for proposed methodology.

### A. Dataset Description

The research study was conducted using a dataset of CT Kidney images, sourced from an official repository accessible [36]. This dataset was compiled from various hospitals in Bangladesh and Dhaka, obtained through the Picture Archiving and Communication System (PACS). The dataset comprises a total of 12,446 images, distributed across different categories: 3,709 images representing cyst, 5,077 images of normal kidney, 1,377 images of kidney stone, and 2,283 images of kidney tumor patients. Each image in the dataset possesses a resolution of 512 x 512 pixels.

The research investigation primarily focused on two classes within the dataset, specifically the normal and tumor images. Fig. 2 present the CT slide of normal and tumorous kidney samples. To facilitate the experimentation process, the dataset was stratified into three distinct subsets, maintaining a

partitioning ratio of 6:2:2, representing 60% for training, 20% for validation, and 20% for testing purposes, respectively. Such a division ensures a robust evaluation of the proposed methods and allows for the assessment of model performance across distinct data subsets.



(a) (b)

Fig. 2. (a) CT slice of a normal kidney (b) CT slice of kidney with tumor.

*1) Data Pre-processing*: This dataset is taken through three distinct data pre-processing stages before being experimented as discussed below:

*a) Thresholding*: It is the simplest type of image segmentation technique which is applied on the gray scale images to convert them to binary images. The coloured images are converted to binary images using thresholding [37]. Thresholding is experimented to divide a picture into less sections, or trashes, utilizing somewhere around one shading or dim scale an incentive to characterize their boundaries. The complexity of the information is simplified, and the process of detecting and characterising the data can be clarified if a binary image is obtained first is one of the advantages of this approach. Selecting a single Threshold value (T) is the method that is the most well-known for transforming a grayscale image into a binary one. Mathematically it can be interpreted as follows:

$$O_{i,j} = \begin{cases} 1 \text{ if } I_{i,j} >= T \\ 0 \text{ if } I_{i,j} < T \end{cases} \quad (1)$$

*b) Data normalization:* The image pixels are normalized to a single scale of 0 to 255 gray scale values.

*c) Image resizing:* The given dataset images of dimension 512 x 512 are resized to (32x 32), (64 x 64), and (128x 128) to perform the experiment.

*2) Feature engineering*: It is the amalgamation of the feature extraction, selection and matrix creation. In the proposed research work, features are extracted from the given input images. The Grey Level Co-occurrence matrix (GLCM) [38] has been generated for all the images for the extraction of the relevant features. The GLCM is calculated to analyse the texture of the given image based upon the contrast, homogeneity, Correlation between the matrix, dissimilarity and energy of the pixels. The GLCM uses the distance(d) and angle (θ) between pair of pixels (i, j) to calculate the occurrence of the (i, j)$^{th}$ pair in the image in the direction of the given angle value. The θ value varies from 0 to 360 degree. The size of the matrix depends upon the number of pixel intensities present in the given image for which the matrix has to be calculated.

For example: If there are four intensity values present in the image , then the size of the GLCM will be 4 x 4. The GLCM is calculated for the three different image dimensions i.e. (32x 32), (64 x 64), and (128x 128). But different image sizes do not have much impact on the classification accuracy. So we consider the smallest image size for the feature extraction and model training in order to avoid the unnecessary computation cost. The process of GLCM is repeated for four times. The final extracted features are then combined and given to the Light Gradient Boosting Machine (LGBM) for training.

### *B. Machine Learning Methods*

*1) Random Forest (RF)*: The RF [39] algorithm is a well-known ensemble learning method that is utilized for classification purpose, including kidney disease classification. Random Forest is applied on kidney disease dataset which is prepared with features and corresponding target labels. The features may include various clinical and demographic factors related to kidney disease, while the target labels indicate the presence or absence of kidney disease.

Random Forest Model training follows the below mentioned steps:

- Tree Construction: RF consists of multiple decision trees (DT). Where each tree is constructed during training by taking a random subset from training data, known as bootstrap samples. This sampling technique introduces variation in the training process.
- Feature Sampling: An unsystematic subset of features is considered at each node of the decision tree, for splitting known as feature sampling or feature bagging. It helps in reducing correlation among the trees and promotes diversity in the ensemble.
- Splitting Criteria: The decision tree nodes are split using a splitting (Gini impurity) criterion that helps determining the best feature and threshold to divide the data at each node, aiming to maximize the separation of the target classes.
- Tree Growth: The trees are grown until a stopping condition is met. Controlling the tree growth helps prevent overfitting and promotes generalization.

*a) Voting Mechanism:* Once the RF model is trained, predictions are made by aggregating the predictions of all the individual decision trees. The final predictions made by each tree in the forest is determined by majority vote obtained independently, and the final prediction is determined by a majority vote or averaging, depending on the task (binary or multi-class classification).

*b) Class Probability Estimation:* RF can also estimate the probability of belonging to each class. It does this by calculating the proportion of trees that predicted each class. The class probabilities can be useful for assessing the confidence of predictions or for other downstream tasks.

The trained RF is evaluated on the testing dataset to assess its performance using metrics that present insights into the

model's prognostic accuracy and capability to correctly classify instances of kidney disease. The trained Random Forest model can be deployed to make predictions on new, unseen data. It takes the relevant features of a patient as input and provides the predicted class or class probabilities for kidney disease. RF offers several advantages for kidney disease classification, including the ability to handle high-dimensional datasets, handle missing values, and provide feature importance measures. It is a robust and widely used algorithm for classification tasks, including medical diagnosis and risk assessment.

*2) Support Vector Machine (SVM)*: (SVM) [40] is supervised learning algorithm commonly used for classification tasks. The internal architecture of an SVM involves several key components and steps.

- Relevant features: Relevant features from the kidney disease dataset are extracted. These features could include measurements such as blood pressure, serum creatinine levels, age, etc.
- Data Normalization: The retrieved characteristics are then normalized such that they are all on a scale that is comparable to one another. This step helps in preventing definite features from dominating the others due to their larger magnitudes.
- Feature Representation: Each instance or sample of dataset is presented as a feature vector, where each feature corresponds to a specific attribute or measurement related to kidney disease.
- Hyperplane Initialization: In a high-dimensional feature space, the SVM finds the best hyper plane to split data points by class. This hyperplane gets configured at the outset by the SVM algorithm.
- Training: The SVM algorithm trains the model by iteratively optimizing the position and orientation of the hyperplane to get the most out of the margin between the classes. The goal is to find the hyperplane that separates the classes by minimizing the misclassification. The optimization problem is typically solved using techniques such as the Sequential Minimal Optimization (SMO) algorithm or other quadratic programming methods. During the training process, subset of data points called support vectors are identifies that are closest to the hyperplane or located inside the boundary.
- Classification: After training is complete, SVM is used for classification of new, unseen instances. The decision boundary of the SVM is defined by the hyperplane, and the side of the hyperplane on which a data point lies determines its predicted class label. The distance of a data point from the hyperplane can also provide additional information about the model's confidence in its prediction.

*3) Gradient Boosting (GB)*: GB [41] is an ensemble learning method that incorporated multiple weak learners, to create a strong prognostic model. GB is used for kidney disease classification is discussed stepwise as below:

- Data Preparation: Similar to other machine learning algorithms, the kidney disease dataset needs to be prepared by extracting relevant features and normalizing the data if necessary.
- Initialization: Gradient Boosting starts with an initial model, often a simple one like a decision tree with limited depth (weak learner). The initial model is typically assigned equal weights for all samples in the training set.
- Iterative Training: In each iteration, a new weak learner, referred to as a "base learner," is trained to approve the miss-classifications made by the ensemble of models trained so far. The base learner is fitted to the training set, with a focus on the samples that were misclassified or had high residuals from the previous iteration. The learning process involves minimizing a loss function. The base learner is typically a decision tree that is grown using a greedy algorithm, selecting the best split points based on information gain or other criteria. The depth and complexity of the decision tree can be adjusted to balance model performance and computational efficiency.
- Boosting and Weight Updates: After training the base learner, its predictions are combined with the predictions of the previous models in the ensemble. Initially, all models are given equal weights. However, the subsequent models focus on the samples that were misclassified or had high residuals from the previous models. The weights of the samples are adjusted to prioritize the challenging instances. The misclassified samples are assigned higher weights, while correctly classified samples have lower weights. This process emphasizes the samples that are difficult to classify, allowing subsequent models to concentrate on improving their predictions.
- Iteration and Ensemble Building: Steps 3 and 4 are repetitive for a predetermined number of iterations or until a certain performance threshold is reached. In each iteration, a new base learner is trained to minimize the weighted loss function. The predictions of all models in the ensemble are combined using a weighted sum or averaging scheme, where the weights are determined by the performance of each model.
- Final Prediction: The final prediction for a new, unseen instance is obtained by aggregating the predictions of all models in the ensemble, typically using a majority vote or weighted average. For classification tasks, the predicted class label is determined based on the aggregated predictions.

*4) Light Gradient Boosting Model (LGBM)*: LGBM [42] is a powerful gradient boosting framework that combines speed and efficiency with high predictive accuracy. Its ability to handle large datasets efficiently and its regularization techniques make it a popular choice among data scientists and

machine learning practitioners. Light GBM is based on the gradient boosting framework, which is an ensemble method. It sequentially trains the models to overcome the drawbacks of the previous models, thus getting better the overall prognostic accuracy.

The kidney disease dataset is prepared with features and corresponding target labels. The features may include various clinical and demographic factors related to kidney disease, such as age, blood pressure, creatinine levels, etc. The target labels indicate the presence or absence of kidney disease. Model Configuration: The LGBM model is configured with the following parameters:

- Learning Rate: The learning rate determines the step size at each boosting iteration. In this case, the learning rate is set to 0.05, indicating a relatively small step size.
- Boost Type: The boost type is set to "Dart," which refers to the Dart boosting algorithm. Dart is a variation of gradient boosting that introduces dropout regularization to prevent overfitting.
- Metric: The chosen evaluation metric is "multi log loss," which is suitable for multi-class classification problems. It calculates the logarithmic loss between the predicted labels and true class labels.
- Number of Leaves: The LGBM model is configured with 100 leaves. The leaves represent the final decision regions of the ensemble of decision trees.
- Max Depth: of each individual decision tree in the ensemble is set to ten. This parameter limits the complexity of the trees and helps prevent overfitting.
- Class: The classification task involves predicting between two classes, likely representing the presence or absence of kidney disease.

The LGBM model is trained on the training dataset using the configured parameters. The model uses gradient boosting to iteratively fit decision trees to the training data, improving its predictive performance at each iteration. Once the training is complete, the trained LGBM model is evaluated on the testing dataset using the multi log loss metric. It can take the relevant features of a patient as input and provide the probability or predicted class of kidney disease. By following this architectural workflow, the LGBM model with the specified parameters can be effectively trained and used for kidney disease classification. It is important to note that further customization and tuning of the parameters might be necessary depending on the specific description of the dataset and the desired performance. The pseudo code for the proposed research methodology is given in Pseudocode 1.

**# 1. Pseudocode for ML models for kidney disease classification**

```
Step 1: Load and preprocess the dataset // Load dataset
   Preprocess dataset (handle missing values, encode categorical variables, etc.)
   Split dataset into:  features (X) and target variable (y)
Step 2: train_test_split is done
Step 3:  Model1 = RFClassifier( ),
   Model2 = GBClassifier(),
    Model3 = SVMClassifier()',
    Define the model "Light GBM"  and set hyperparameters
     Model4 = LGBMClassifier( boosting_type='dart',
     objective='binary',
     metric='multi logloss', num_leaves = <100>,
     learning_rate = <0.05>,
     feature_fraction = <feature_fraction>,
     bagging_fraction = <bagging_fraction>,
    verbose = -1)
  Step 4: model.fit(X_train, y_train) //Train the Light GBM model
  Step 5: y_pred =  model.predict(X_test) //Make predictions and evaluate the model
  Step 6: Accuracy = accuracy_score(y_test, y_pred) // Calculate evaluation metrics
        CM = confusion_matrix(y_test, y_pred)
  Step 7: Tune hyperparameters to optimize model performance
Step 8: Perform further analysis (e.g., feature importance)
```

LGBM has shown excellent prediction for kidney disease classification. Its ability to iteratively refine the model by focusing on the challenging instances makes it effective in capturing complex relationships in the data and improving predictive accuracy.

### *C. Transfer Learning Architectures*

*1) ResNet-101*: ResNet-101 [43] features a 101-layer deep CNN. The weights of the network's pre-trained version on over a million photos from the ImageNet database may be imported. The trained network can classify photos into 1000 different things, such as keyboards, mice, books, and other stuff. To train the model, an Adam optimizer with a learning rate of 4e-5 and binary cross entropy as the loss function was used. During model training, the total number of trainable parameters was 525,313.

*2) DenseNet-121*: It is a dense network [44] of 121 layers, 120 of which are dense layers and four of which are average pool layers. The weights of all layers in the same deep dense block are circulated across the inputs, allowing the deep layers to utilise the early extracted features. DenseNet uses features more efficiently and outperforms with less parameters. The network is trained over ten epochs using an Adam optimizer and a learning rate of 4e-5 hyperparameters. The total number of trainable parameters in the model was 263,169.

The internal architecture for fine-tuned ResNet-101 and DenseNet-121 are given in Fig. 3 and 4, respectively. The accuracy of the model was found to be enhanced when (32, 32) pixel images were used as input. The smaller image size results in a lower computation and time complexity, which may cause the model to learn features or patterns in the images faster. The Adam optimizer was used to train the model since it is the best optimizer for early convergence. When employing Adam as an optimizer, the model learns up new information more quickly. In order to train our suggested model, we implement a learning rate schedule in the Adam. The training of the model lasted 50 epochs, with a batch size of 32, and the initial learning rate was set at 0.001.

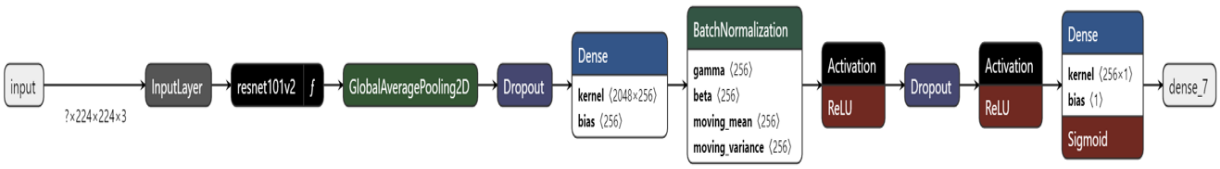



Fig. 3. Architecture of ResNet-101 model for diagnosing KT using CT images.

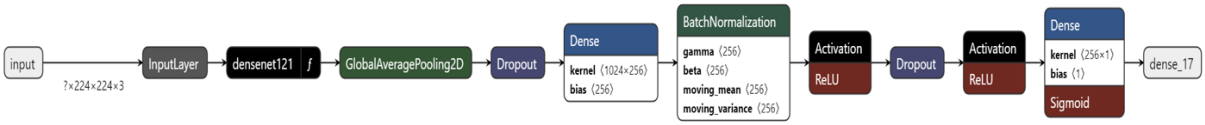



Fig. 4. Architecture of DenseNet-121 model for diagnosing KT using CT images.

### D. Experimental Environment Settings and Performance Evaluation Metrics

This research aims to propose an optimal model which identifies and classifies different types of images. The proposed model was implemented using Python (v. 3.8), OpenCV (v. 4.7), Keras Library (v. 2.8) were used on Windows 10 Pro OS, with system configuration using an Intel i5 processor running at 2.9 GHz, an Nvidia RTX 2060 Graphical Processing Unit and 16 GB RAM.

Several metrics were employed to evaluate the performance of classifying sunflower blooms and leaves, including accuracy, precision, recall, and F1-score, which are frequently used indicators [45] . Accuracy is the ratio of samples from all classes that can be correctly identified, Recall is the ratio of correctly classified positives among all actual positives, and Precision is the ratio of correctly identified positives versus all expected positives [46]. The metrics were calculated using Eq. (1) - (4).

$$Accuracy = \frac{True\ Positive+True\ Negative}{True\ Positive+True\ Negative+False\ Positive+False\ Negative} \quad (1)$$

$$Recall = \frac{True\ Positive}{True\ Positive+False\ Negative} \quad (2)$$

$$Precision = \frac{TP}{True\ Positive+False\ Positive} \quad (3)$$

$$F1-Score = 2x\frac{RecallxPrecision}{Recall+Precision} \quad (4)$$

$$AUC = \sum_{i=0}^{N}\left(\frac{TPI(i)+TPI\ (i-1)}{2}\right)(FPI\ (i)-(FPI\ (i-1)) \quad (5)$$

## III. Results and Discussion

### A. Performance Comparison

The ML models were meticulously trained on a dataset, and to bolster their performance, advanced feature extraction using the GLCM technique was employed. This technique allowed the models to capture intricate patterns and textures from the kidney images, enhancing their ability to predict Kidney Tumors (KT) more accurately. To push the boundaries of prediction accuracy even further, two state-of-the-art architectures, ResNet-101 and DenseNet-121, were extensively experimented with. These cutting-edge models were chosen for their exceptional deep learning capabilities, enabling them to unravel complex relationships within the data and make precise predictions regarding the presence of kidney tumors.

Following the intensive training process, the models' prowess was thoroughly evaluated using a diverse range of performance metrics, such as accuracy, recall, precision, and F1_score. These metrics provided a comprehensive view of the models' overall classification performance and offered valuable insights into their strengths and potential areas of improvement. In addition to the conventional performance metrics, the proposed model's efficacy was also measured using the Area Under the Curve (AUC) on the test set. AUC is a significant indicator of the model's ability to distinguish between positive and negative cases, providing a more comprehensive understanding of its discriminative power.

The outcomes of the classification process were further analyzed to gain a deeper understanding of the models' predictions. Specifically, the binary class results (class 0 and class 1) were meticulously generated and assessed. Label 0 corresponded to the classification results for normal kidneys, indicating instances where the model correctly identified normal kidneys (True negative). On the other hand, Label 1 represented the classification results for tumorous kidneys, showcasing the model's capacity to accurately detect kidney tumors (True positive).

Moreover, the results also provided insights into false positives and false negatives, highlighting instances where the model might have made errors in its predictions. This comprehensive analysis aimed to identify potential areas of improvement and guide future iterations of the models. By combining advanced feature extraction techniques, cutting-edge deep learning architectures, and a thorough evaluation using diverse performance metrics, this study aimed to achieve a robust and reliable prediction model for Kidney Tumor detection, which could have a significant impact on the early diagnosis and treatment of kidney diseases.

Table I presents the results of various performance metrics calculated on the Test set for different machine learning algorithms, as well as for the proposed fine-tuned Transfer Learning architectures, ResNet-101, and DenseNet-121. The metrics evaluated are Accuracy, Precision, Recall, F1_score, and AUC.

TABLE I. RESULTS OF THE DIFFERENT PERFORMANCE METRICS CALCULATED ON THE TEST SET

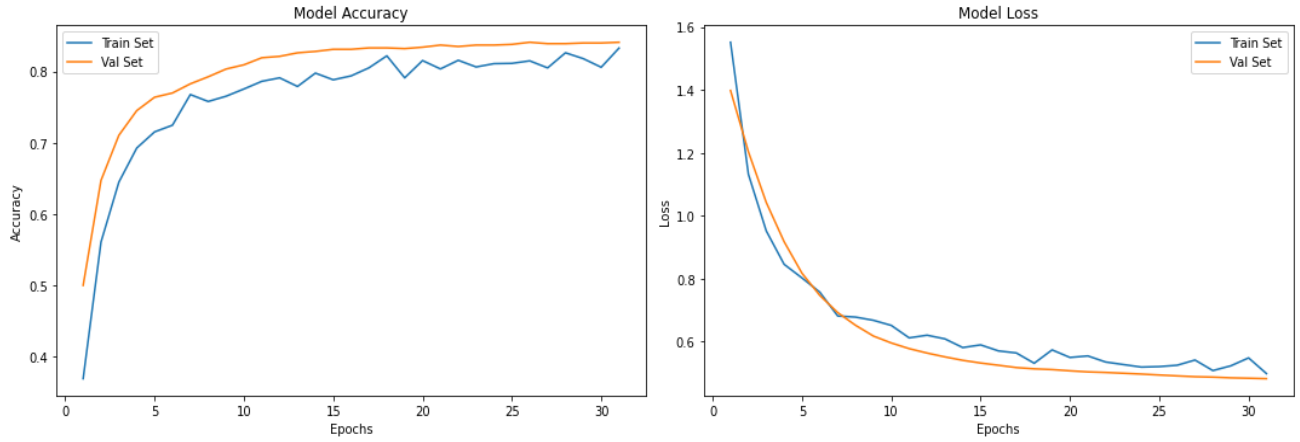
| Model | Accuracy | Precision | Recall | F1_score | AUC |
|---|---|---|---|---|---|
| LGBM | 94.09 | 0.9510 | 0.9352 | 0.9595 | 0.9552 |
| GB | 92.44 | 0.9439 | 0.8991 | 0.9843 | 0.9773 |
| SVM | 91.27 | 0.9300 | 0.8890 | 0.8906 | 0.8834 |
| RF | 91.02 | 0.9201 | 0.9212 | 0.9656 | 0.9588 |
| ResNet-101 | 96.67 | 0.9532 | 0.9111 | 0.9843 | 0.9773 |
| DenseNet-121 | 98.22 | 0.9577 | 0.9323 | 0.9843 | 0.9773 |

From the table it can be noticed that both ResNet-101 and DenseNet-121 achieved the highest accuracy among all the models. DenseNet-121 outperformed all other models with an impressive accuracy of 98.22%, while ResNet-101 achieved an accuracy of 96.67%. The success of these Transfer Learning architectures can be attributed to their pre-trained weights and knowledge gained from large datasets, allowing them to recognize and learn intricate patterns and features from the given kidney tumor dataset effectively. LGBM and GB achieved reasonably high accuracies of 94.09% and 92.44%, respectively. LGBM performed slightly better than GB, which indicates the effectiveness of gradient boosting in ensemble learning. However, the accuracies of both LGBM and GB were lower compared to the Transfer Learning models. SVM and RF achieved accuracies of 91.27% and 91.02%, respectively. While SVM relies on finding optimal hyperplanes for classification, RF uses an ensemble of decision trees. Although these algorithms achieved respectable accuracies, they were outperformed by the Transfer Learning models. The precision, recall, and F1_score metrics provide insights into the models' ability to correctly classify positive and negative instances, as well as their overall predictive performance. DenseNet-121 consistently achieved the highest F1_score of 0.9843, indicating its superior balance between precision and recall in classifying both tumor and normal kidney instances. AUC is a measure of the models' ability to distinguish between positive and negative cases. Remarkably, both ResNet-101 and DenseNet-121 attained an AUC of 0.9773, matching the performance of GB. This demonstrates the Transfer Learning models' robustness in making accurate predictions and differentiating between tumor and normal kidney instances.

Accuracy is a crucial metric in evaluating the performance of machine learning models, as it represents the percentage of correct predictions made by the model where the predicted value aligns with the real value. Throughout the training phase, accuracy is continuously monitored and plotted, providing valuable insights into the model's learning progress and its ability to make accurate predictions as it iteratively updates its weights and biases.

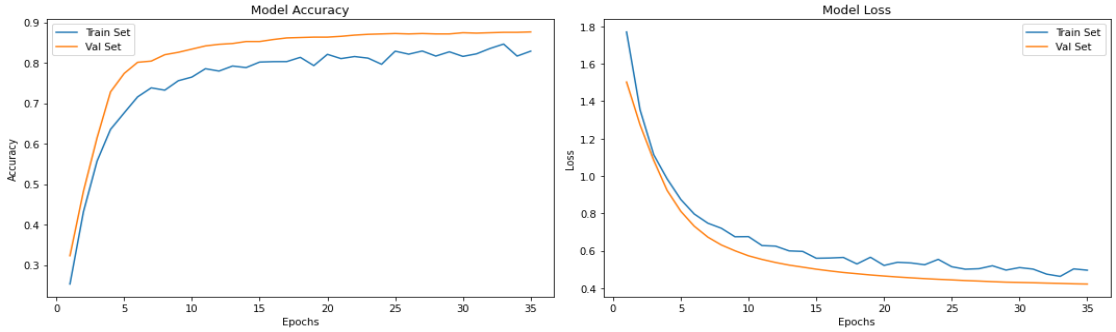
While accuracy provides an overall view of the model's correctness, it is essential to delve deeper into the model's learning dynamics. For this purpose, loss functions play a pivotal role in assessing the model's performance. Loss functions measure the disparities between predicted values and actual ground-truth values, quantifying the uncertainty or error in the model's estimates. By optimizing the loss during training, the model learns to minimize the discrepancies and improve its predictive capability.

In Fig. 5 and Fig. 6, we present the accuracy and loss plots for two cutting-edge deep learning architectures, ResNet-101 and DenseNet-121, respectively. These plots showcase how accuracy improves and loss decreases over the training iterations, providing a comprehensive view of the models' learning behaviors. The ascending accuracy curve demonstrates how the models become more adept at correctly classifying kidney tumor images as training progresses. Simultaneously, the descending loss curve indicates that the models effectively minimize prediction errors, leading to more precise and confident predictions. Analyzing the accuracy and loss plots for ResNet-101 and DenseNet-121 offers valuable insights into their learning dynamics and convergence patterns. These visualizations not only validate the models' effectiveness in the classification task but also aid in fine-tuning the hyperparameters or adjusting the training strategy to achieve optimal performance. By carefully monitoring the accuracy and loss during training, we gain a deeper understanding of the models' efficacy and can confidently assert that both ResNet-101 and DenseNet-121 exhibit exceptional learning capabilities, making them powerful tools for kidney tumor classification.




(a) (b)

Fig. 5. Graphs plotted between model (a) accuracy and (b) loss over no of epochs for ResNet-101.

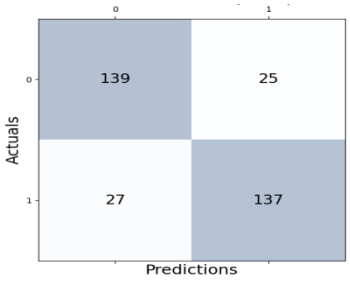



(a) (b)

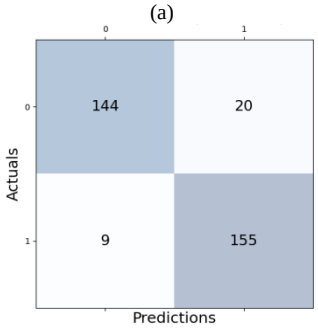
Fig. 6. Graphs plotted between model (a) accuracy and (b) loss over no of epochs DenseNet-121.

In addition to accuracy and loss analysis, the confusion matrix (CM) stands as a pivotal tool for the comprehensive evaluation of classification models, such as ResNet-101 and DenseNet-121. The CM provides valuable insights into the model's performance by breaking down the predictions into four fundamental categories: true positives, true negatives, false positives, and false negatives. It highlights the model's ability to correctly classify positive and negative instances, as well as its potential for making erroneous predictions.

Fig. 7(a) and 7(b) depict the confusion matrices specifically for ResNet-101 and DenseNet-121, respectively. These visual representations offer a detailed view of the models' classification outcomes, enabling us to gauge their effectiveness in distinguishing between normal and tumorous kidney images.




(a)




(b)

Fig. 7. Confusion Matrix of (a) ResNet-101 and (b) DenseNet-121.

## B. K-Fold Cross Validation

K-fold cross-validation is a widely used technique in machine learning to assess the performance and generalization capability of a model while mitigating the risks of overfitting. Overfitting occurs when a model performs well on the training data but fails to generalize to unseen data, which can lead to inflated evaluation metrics on the test set. To address this concern, the model's performance is tested using a separate validation set, and 10-fold cross-validation is a popular approach to achieve this.

In 10-fold cross-validation, the dataset is divided into 10 subsets of approximately equal size. The model is then trained and evaluated ten times, each time using a different subset as the validation set and the remaining nine subsets for training. This ensures that the model is evaluated on different partitions of the data, providing a more robust estimate of its performance and reducing the influence of any particular data split.

Table II presents the results of 10-fold cross-validation on the CT Kidney image dataset for different machine learning (ML) models and fine-tuned Transfer Learning (TL) models, namely Random Forest, Support Vector Machine (SVM), Gradient Boost, LGBM (LightGBM), ResNet-101, and DenseNet-121. The accuracies achieved by each model on each fold are shown.

Observations from the 10-fold Cross Validation Results:

*1)* Random Forest: Random Forest demonstrates relatively stable performance across the folds, with accuracy ranging from 70.12% to 84.21%. Its mean accuracy is 77.09%.

*2)* Support Vector Machine (SVM): SVM shows higher accuracy values, with a range of 66.23% to 81.81% across the folds. The mean accuracy is 88.82%, making it one of the best-performing ML models in this study.

*3)* Gradient Boost: Gradient Boosting performs consistently well, with accuracy varying between 62.33% to 79.12%. The mean accuracy achieved by Gradient Boost is 89.11%.

*4)* LGBM (LightGBM): LGBM outperforms other ML models, exhibiting accuracy in the range of 89.12% to 93.32%. Its mean accuracy is 90.71%.

*5)* ResNet-101 and DenseNet-121 (Transfer Learning): Both fine-tuned Transfer Learning models, ResNet-101, and DenseNet-121 consistently achieve higher accuracy values compared to ML models. ResNet-101 achieves accuracy between 86.72% to 96.81%, with a mean accuracy of 91.61%. DenseNet-121 exhibits even better performance, with accuracy ranging from 90.0% to 97.61% and a mean accuracy of 92.44%.

TABLE II. 10 FOLD CROSS VALIDATION PERFORMANCE OF THE ML AND FINE-TUNED TL MODELS ON THE CT KIDNEY IMAGE DATASET

| 10-folds | Random Forest | Support Vector Machine | Gradient Boost | LGBM | ResNet-101 | DensNet-121 |
|---|---|---|---|---|---|---|
| Fold-1 | 79.22 | 81.81 | 74.02 | 91.62 | 88.92 | 96.81 |
| Fold-2 | 70.12 | 66.23 | 70.12 | 89.12 | 86.72 | 92.02 |
| Fold-3 | 80.51 | 79.22 | 72.83 | 88.92 | 96.62 | 94.62 |
| Fold-4 | 79.22 | 75.32 | 79.12 | 89.23 | 92.62 | 94.82 |
| Fold-5 | 72.72 | 71.42 | 77.42 | 91.92 | 91.02 | 91.72 |
| Fold-6 | 81.81 | 77.92 | 79.12 | 92.32 | 90.02 | 92.82 |
| Fold-7 | 71.42 | 68.83 | 62.33 | 93.23 | 92.33 | 87.32 |
| Fold-8 | 75.32 | 76.62 | 78.42 | 91.83 | 91.62 | 90.32 |
| Fold-9 | 76.31 | 76.31 | 71.05 | 93.32 | 90.0 | 91.71 |
| Fold-10 | 84.21 | 81.57 | 77.63 | 89.57 | 90.26 | 97.61 |
| Mean | 77.09 | 88.82 | 89.11 | 90.71 | 91.61 | 92.44 |

Overall, the results demonstrate the effectiveness of the fine-tuned Transfer Learning models, ResNet-101 and DenseNet-121, in accurately classifying kidney images. These models outperform the traditional ML algorithms, such as Random Forest, SVM, and Gradient Boost, in terms of accuracy. The use of 10-fold cross-validation provides a more reliable estimate of the models' performance and their generalization capability, ensuring that the evaluations are robust and less affected by variations in the data splits.

In conclusion, the 10-fold cross-validation results reveal the superior performance of the fine-tuned Transfer Learning models, ResNet-101, and DenseNet-121, in accurately classifying CT kidney images. These models are better equipped to handle the complexities of the dataset, offering promising implications for kidney tumor detection and diagnosis in a real-world clinical setting.

Furthermore, Table III presents a comparison of the results obtained using LightGBM (LGBM) with state-of-the-art methods in terms of accuracy for a specific task or dataset. The table showcases the performance of various models, including two fine-tuned Transfer Learning architectures, ResNet-101 and DenseNet-121, along with several other approaches reported in the literature.

Table III demonstrate the effectiveness of the fine-tuned Transfer Learning models, ResNet-101 and DenseNet-121, in achieving high accuracies and outperforming other state-of-the-art methods. DenseNet-121, in particular, exhibits the highest accuracy among all the models, indicating its superior performance in the specific classification task. The results also highlight the importance of exploring and comparing different models and methodologies to advance the field and achieve better results in image classification and other related tasks.

TABLE III. COMPARISON OF PROPOSED WORK WITH STATE-OF-THE-ART METHODS

| Reference | Accuracy (%) |
|---|---|
| ResNet-101 | 96.67 |
| DenseNet-121 | 98.22 |
| LGBM | 94.09 |
| Zhou et al.[47] | 93.00 |
| Zabihollahy et al. [48] | 83.75 |
| Schieda et al. [49] | 78.00 |
| Finally, Yap et al. [50] | 75.00 |

## IV. CONCLUSION

The early detection and classification of kidney tumors play a vital role in saving human lives. Manual detection methods rely heavily on the expertise of medical professionals and can be time-consuming. Therefore, the development of automatic classification systems holds significant promise, as they offer robust and rapid results. In this study, we have presented a hybrid approach combining Light Gradient Boosting Method with Grey Level Co-occurrence matrix (GLCM) computation for the automatic classification of kidney tumors from CT Kidney image datasets. To optimize the training process, we applied various pre-processing techniques and image resizing, reducing the model's complexity and speeding up the training. Light GBM, known for its speed, efficiency, and high predictive accuracy among ML models, served as a powerful gradient boosting framework in this study. Additionally, we introduced two fine-tuned Transfer Learning (TL) models within this framework, ResNet-101 and DenseNet-121, to predict kidney tumors. The performance of these models was thoroughly evaluated using diverse performance metrics and compared with state-of-the-art methods. Our results demonstrated the superiority of the fine-tuned TL models, with DenseNet-121 achieving an impressive accuracy of 98.22%.

Several limitations are evident in the study's comparative analysis, model testing, and generalizability. Firstly, the comparative study of detection approaches, including Random Forest, Support Vector Machine, Gradient Boosting, Light Gradient Boosting Model, and deep learning models ResNet-101 and DenseNet-121, might lack a comprehensive exploration of other relevant models, potentially missing out on valuable insights and alternative solutions. Secondly, while the fine-tuned deep learning models exhibit impressive accuracy, their testing and evaluation solely on the provided dataset might not guarantee similar performance on diverse and real-world datasets. There's a need to assess the models across various datasets to ascertain their consistency and robustness. Finally, the proposed approach's applicability to different datasets with varying characteristics, like imaging quality and patient demographics, remains unexplored. Testing the models on multiple datasets could reveal potential challenges in generalizing the approach to broader clinical settings.

In light of these limitations, future work should aim to address these areas to enhance the study's comprehensiveness and practicality. To conduct a more thorough comparative study, incorporating a wider range of detection models, including emerging techniques and architectures, would provide a more comprehensive understanding of the strengths and weaknesses of various approaches. Moreover, the robustness and generalizability of the fine-tuned deep learning models, ResNet-101 and DenseNet-121, should be evaluated across multiple datasets to ensure consistent performance across diverse clinical scenarios. Additionally, to enhance the proposed approach's real-world applicability, further investigation on different datasets, encompassing variations in imaging quality, patient populations, and demographics, is crucial. This analysis will shed light on potential challenges and adaptations needed to deploy the model effectively in clinical practice. Ultimately, addressing these avenues for future research will contribute to a more holistic and adaptable approach for kidney tumor detection, ensuring its utility and effectiveness across a broader spectrum of clinical settings.

## ACKNOWLEDGMENT

King Faisal University: Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia, under the Project GRANT3, 926.

## REFERENCES

[1] Moch, H.; Cubilla, A.L.; Humphrey, P.A.; Reuter, V.E.; Ulbright, T.M. The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs—Part A: Renal, Penile, and Testicular Tumours. Eur Urol 2016, 70, doi:10.1016/j.eururo.2016.02.029.

[2] Rowe, S.P.; Pomper, M.G. Molecular Imaging in Oncology: Current Impact and Future Directions. CA Cancer J Clin 2022, 72, 333–352, doi:10.3322/caac.21713.

[3] Ansari, K.K.; Jha, A. Causes of Cancer in the World: Comparative Risk Assessment of Nine Behavioral and Environmental Risk Factors. Cureus 2022, 14, e28875–e28875, doi:10.7759/cureus.28875.

[4] Kidney Cancer: Statistics Available online: Kidney Cancer: Statistics (accessed on 23 July 2023).

[5] Khan, Y.F.; Kaushik, B. Neuroimaging (Anatomical MRI)-Based Classification of Alzheimer's Diseases and Mild Cognitive Impairment Using Convolution Neural Network. In Lecture Notes on Data Engineering and Communications Technologies; 2022; Vol. 106.

[6] Irigaray, P.; Newby, J.A.; Clapp, R.; Hardell, L.; Howard, V.; Montagnier, L.; Epstein, S.; Belpomme, D. Lifestyle-Related Factors and Environmental Agents Causing Cancer: An Overview. Biomedicine & Pharmacotherapy 2007, 61, 640–658, doi:10.1016/j.biopha.2007.10.006.

[7] Kocak, B.; Durmaz, E.S.; Kaya, O.K.; Ates, E.; Kilickesmez, O. Reliability of Single-Slice–Based 2D CT Texture Analysis of Renal Masses: Influence of Intra- and Interobserver Manual Segmentation Variability on Radiomic Feature Reproducibility. American Journal of Roentgenology 2019, 213, 377–383, doi:10.2214/ajr.19.21212.

[8] Gulzar, Y. Fruit Image Classification Model Based on MobileNetV2 with Deep Transfer Learning Technique. Sustainability 2023, 15, 1906.

[9] Dhiman, P.; Kaur, A.; Balasaraswathi, V.R.; Gulzar, Y.; Alwan, A.A.; Hamid, Y. Image Acquisition, Preprocessing and Classification of Citrus Fruit Diseases: A Systematic Literature Review. Sustainability 2023, Vol. 15, Page 9643 2023, 15, 9643, doi:10.3390/SU15129643.

[10] Mamat, N.; Othman, M.F.; Abdulghafor, R.; Alwan, A.A.; Gulzar, Y. Enhancing Image Annotation Technique of Fruit Classification Using a Deep Learning Approach. Sustainability 2023, 15, 901.

[11] Gulzar, Y.; Ünal, Z.; Akta¸s, H.A.; Mir, M.S. Harnessing the Power of Transfer Learning in Sunflower Disease Detection: A Comparative Study. Agriculture 2023, Vol. 13, Page 1479 2023, 13, 1479, doi:10.3390/AGRICULTURE13081479.

[12] Aggarwal, S.; Gupta, S.; Gupta, D.; Gulzar, Y.; Juneja, S.; Alwan, A.A.; Nauman, A. An Artificial Intelligence-Based Stacked Ensemble Approach for Prediction of Protein Subcellular Localization in Confocal Microscopy Images. Sustainability 2023, Vol. 15, Page 1695 2023, 15, 1695, doi:10.3390/SU15021695.

[13] Gulzar, Y.; Hamid, Y.; Soomro, A.B.; Alwan, A.A.; Journaux, L. A Convolution Neural Network-Based Seed Classification System. Symmetry (Basel) 2020, 12, 2018.

[14] Sahlan, F.; Hamidi, F.; Misrat, M.Z.; Adli, M.H.; Wani, S.; Gulzar, Y. Prediction of Mental Health Among University Students. International Journal on Perceptive and Cognitive Computing 2021, 7, 85–91.

[15] Hamid, Y.; Elyassami, S.; Gulzar, Y.; Balasaraswathi, V.R.; Habuza, T.; Wani, S. An Improvised CNN Model for Fake Image Detection. International Journal of Information Technology 2022 2022, 1–11, doi:10.1007/S41870-022-01130-5.

[16] Gulzar, Y.; Alwan, A.A.; Abdullah, R.M.; Abualkishik, A.Z.; Oumrani, M. OCA: Ordered Clustering-Based Algorithm for E-Commerce Recommendation System. Sustainability 2023, Vol. 15, Page 2947 2023, 15, 2947, doi:10.3390/SU15042947.

[17] Gulzar, Y.; Khan, S.A. Skin Lesion Segmentation Based on Vision Transformers and Convolutional Neural Networks—A Comparative Study. Applied Sciences 2022, Vol. 12, Page 5990 2022, 12, 5990, doi:10.3390/APP12125990.

[18] Khan, S.A.; Gulzar, Y.; Turaev, S.; Peng, Y.S. A Modified HSIFT Descriptor for Medical Image Classification of Anatomy Objects. Symmetry (Basel) 2021, 13, 1987.

[19] Alam, S.; Raja, P.; Gulzar, Y. Investigation of Machine Learning Methods for Early Prediction of Neurodevelopmental Disorders in Children. Wirel Commun Mob Comput 2022, 2022.

[20] Mehmood, A.; Gulzar, Y.; Ilyas, Q.M.; Jabbari, A.; Ahmad, M.; Iqbal, S. SBXception: A Shallower and Broader Xception Architecture for Efficient Classification of Skin Lesions. Cancers 2023, Vol. 15, Page 3604 2023, 15, 3604, doi:10.3390/CANCERS15143604.

[21] Anand, V.; Gupta, S.; Gupta, D.; Gulzar, Y.; Xin, Q.; Juneja, S.; Shah, A.; Shaikh, A. Weighted Average Ensemble Deep Learning Model for Stratification of Brain Tumor in MRI Images. Diagnostics 2023, Vol. 13, Page 1320 2023, 13, 1320, doi:10.3390/DIAGNOSTICS13071320.

[22] Lee, H.; Hong, H.; Kim, J.; Jung, D.C. Deep Feature Classification of Angiomyolipoma without Visible Fat and Renal Cell Carcinoma in Abdominal Contrast-Enhanced CT Images with Texture Image Patches and Hand-Crafted Feature Concatenation. Med Phys 2018, 45, 1550–1561, doi:10.1002/mp.12828.

[23] Han, S.; Hwang, S.I.; Lee, H.J. The Classification of Renal Cancer in 3-Phase CT Images Using a Deep Learning Method. J Digit Imaging 2019, 32, 638–643, doi:10.1007/s10278-019-00230-2.

[24] Tabibu, S.; Vinod, P.K.; Jawahar, C.V. Pan-Renal Cell Carcinoma Classification and Survival Prediction from Histopathology Images Using Deep Learning. Sci Rep 2019, 9, doi:10.1038/s41598-019-46718-3.

[25] Oberai, A.; Varghese, B.; Cen, S.; Angelini, T.; Hwang, D.; Gill, I.; Aron, M.; Lau, C.; Duddalwar, V. Deep Learning Based Classification of Solid Lipid-Poor Contrast Enhancing Renal Masses Using Contrast Enhanced CT. British Journal of Radiology 2020, 93, doi:10.1259/bjr.20200002.

[26] Pedersen, M.; Andersen, M.B.; Christiansen, H.; Azawi, N.H. Classification of Renal Tumour Using Convolutional Neural Networks to Detect Oncocytoma. Eur J Radiol 2020, 133, doi:10.1016/j.ejrad.2020.109343.

[27] Sudharson, S.; Kokil, P. An Ensemble of Deep Neural Networks for Kidney Ultrasound Image Classification. Comput Methods Programs Biomed 2020, 197, doi:10.1016/j.cmpb.2020.105709.

[28] Pirmoradi, S.; Teshnehlab, M.; Zarghami, N.; Sharifi, A. A Self-Organizing Deep Neuro-Fuzzy System Approach for Classification of Kidney Cancer Subtypes Using MiRNA Genomics Data. Comput

Methods Programs Biomed 2021, 206, doi:10.1016/j.cmpb.2021.106132.

[29] Abdeltawab, H.; Khalifa, F.; Mohammed, M.; Cheng, L.; Gondim, D.; El-Baz, A. A Pyramidal Deep Learning Pipeline for Kidney Whole-Slide Histology Images Classification. Sci Rep 2021, 11, doi:10.1038/s41598-021-99735-6.

[30] Abdeltawab, H.A.; Khalifa, F.A.; Ghazal, M.A.; Cheng, L.; El-Baz, A.S.; Gondim, D.D. A Deep Learning Framework for Automated Classification of Histopathological Kidney Whole-Slide Images. J Pathol Inform 2022, 13, doi:10.1016/j.jpi.2022.100093.

[31] Khan, F.; Ayoub, S.; Gulzar, Y.; Majid, M.; Reegu, F.A.; Mir, M.S.; Soomro, A.B.; Elwasila, O. MRI-Based Effective Ensemble Frameworks for Predicting Human Brain Tumor. Journal of Imaging 2023, Vol. 9, Page 163 2023, 9, 163, doi:10.3390/JIMAGING9080163.

[32] Zhu, X.-L.; Shen, H.-B.; Sun, H.; Duan, L.-X.; Xu, Y.-Y. Improving Segmentation and Classification of Renal Tumors in Small Sample 3D CT Images Using Transfer Learning with Convolutional Neural Networks. Int J Comput Assist Radiol Surg 2022, 17, 1303–1311, doi:10.1007/s11548-022-02587-2.

[33] Gulzar, Y.; Alkinani, A.; Alwan, A.A.; Mehmood, A. Abdomen Fat and Liver Segmentation of CT Scan Images for Determining Obesity and Fatty Liver Correlation. Applied Sciences 2022, Vol. 12, Page 10334 2022, 12, 10334, doi:10.3390/APP122010334.

[34] Sarada, J.; Lakshmi, N.V.M.; Praveena, T.L. U-Capkidnets++-: A Novel Hybrid Capsule Networks with Optimized Deep Feed Forward Networks for an Effective Classification of Kidney Tumours Using CT Kidney Images. International Journal on Recent and Innovation Trends in Computing and Communication 2022, 10, 274–283, doi:10.17762/ijritcc.v10i1s.5849.

[35] Zhao, T.; Sun, Z.; Guo, Y.; Sun, Y.; Zhang, Y.; Wang, X. Automatic Renal Mass Segmentation and Classification on CT Images Based on 3D U-Net and ResNet Algorithms. Front Oncol 2023, 13, doi:10.3389/fonc.2023.1169922.

[36] T KIDNEY DATASET: Normal-Cyst-Tumor and Stone Available online: https://www.kaggle.com/datasets/nazmul0087/ct-kidney-dataset-normal-cyst-tumor-and-stone (accessed on 24 June 2023).

[37] Sahoo, P.K.; Soltani, S.; Wong, A.K.C. A Survey of Thresholding Techniques. Comput Vis Graph Image Process 1988, 41, 233–260, doi:10.1016/0734-189x(88)90022-9.

[38] V, B.S. Grey Level Co-Occurrence Matrices: Generalisation and Some New Features. International Journal of Computer Science, Engineering and Information Technology 2012, 2, 151–157, doi:10.5121/ijcseit.2012.2213.

[39] Breiman, L. Random Forests. Mach Learn 2001, 45, 5–32, doi:10.1023/A:1010933404324/METRICS.

[40] Vapnik, V.; Golowich, S.E.; Smola, A. Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing. In Proceedings of the Advances in Neural Information Processing Systems; 1997.

[41] Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. Ann Stat 2001, 29, doi:10.1214/aos/1013203451.

[42] Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. Adv Neural Inf Process Syst 2017, 30.

[43] He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition; 2016; pp. 770–778.

[44] Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition; 2017; pp. 4700–4708.

[45] Ayoub, S.; Gulzar, Y.; Reegu, F.A.; Turaev, S. Generating Image Captions Using Bahdanau Attention Mechanism and Transfer Learning. Symmetry (Basel) 2022, 14, 2681.

[46] Ayoub, S.; Gulzar, Y.; Rustamov, J.; Jabbari, A.; Reegu, F.A.; Turaev, S. Adversarial Approaches to Tackle Imbalanced Data in Machine Learning. Sustainability 2023, Vol. 15, Page 7097 2023, 15, 7097, doi:10.3390/SU15097097.

[47] Zhou, L.; Zhang, Z.; Chen, Y.-C.; Zhao, Z.-Y.; Yin, X.-D.; Jiang, H.-B. A Deep Learning-Based Radiomics Model for Differentiating Benign and Malignant Renal Tumors. Transl Oncol 2019, 12, 292–300, doi:10.1016/j.tranon.2018.10.012.

[48] Zabihollahy, F.; Schieda, N.; Krishna, S.; Ukwatta, E. Automated Classification of Solid Renal Masses on Contrast-Enhanced Computed Tomography Images Using Convolutional Neural Network with Decision Fusion. Eur Radiol 2020, 30, 5183–5190, doi:10.1007/s00330-020-06787-9.

[49] Schieda, N.; Nguyen, K.; Thornhill, R.E.; McInnes, M.D.F.; Wu, M.; James, N. Importance of Phase Enhancement for Machine Learning Classification of Solid Renal Masses Using Texture Analysis Features at Multi-Phasic CT. Abdominal Radiology 2020, 45, 2786–2796, doi:10.1007/s00261-020-02632-1.

[50] Yap, F.Y.; Varghese, B.A.; Cen, S.Y.; Hwang, D.H.; Lei, X.; Desai, B.; Lau, C.; Yang, L.L.; Fullenkamp, A.J.; Hajian, S.; et al. Shape and Texture-Based Radiomics Signature on CT Effectively Discriminates Benign from Malignant Renal Masses. Eur Radiol 2020, 31, 1011–1021, doi:10.1007/s00330-020-07158-0.