Computers in Biology and Medicine 191 (2025) 110214

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Multi-level feature fusion network for kidney disease detection

Saif Ur Rehman Khan

School of Computer Science and Engineering, Central South University, 932 Lushan S Rd, Yuelu District, Changsha, Hunan, China

ARTICLE INFO

Keywords:
Shap visualization
Feature fusion
Inception block
Recurrent neural network

ABSTRACT

Kidney irregularities pose a significant public health challenge, often leading to severe complications, yet the limited availability of nephrologists makes early detection costly and time-consuming. To address this issue, we propose a deep learning framework for automated kidney disease detection, leveraging feature fusion and sequential modeling techniques to enhance diagnostic accuracy. Our study thoroughly evaluates six pretrained models under identical experimental conditions, identifying ResNet50 and VGG19 as the highly efficient models for feature extraction due to their deep residual learning and hierarchical representations. Our proposed methodology integrates feature fusion with an inception block to extract diverse feature representations while maintaining imbalance dataset overhead. To enhance sequential learning and capture long-term dependencies in disease progression, ConvLSTM is incorporated after feature fusion. Additionally, Inception block is employed after ConvLSTM to refine hierarchical feature extraction, further strengthening the proposed model ability to leverage both spatial and temporal patterns. To validate our approach, we introduce a new named Multiple Hospital Collected (MHC-CT) dataset, consisting of 1860 tumor and 1024 normal kidney CT scans, meticulously annotated by medical experts. Our model achieves 99.60 % accuracy on this dataset, demonstrating its robustness in binary classification. Furthermore, to assess its generalization capability, we evaluate the model on a publicly available benchmark multiclass CT scan dataset, achieving 91.31 % accuracy. The superior performance is attributed to the effective feature fusion using inception blocks and the sequential learning capabilities of ConvLSTM, which together enhance spatial and temporal feature representations. These results highlight the efficacy of the proposed framework in automating kidney disease detection, providing a reliable, and efficient solution for clinical decision-making. https://github.com/VS-EYE/KidneyDiseaseDetection.git.

1. Introduction

Kidney disease is known as renal disease or nephropathy which involves damage to or disease of the kidneys. This impairs their ability to filter blood and remove waste and excess fluids from the body. If left untreated, it can progress to kidney failure, known as end-stage kidney disease, which requires dialysis or a kidney transplant as treatment. Chronic kidney disease-CKD is characterized by a gradual loss of kidney function over time [1]. Common kidney abnormalities include kidney stones, kidney cysts, and kidney tumors, which can be benign or malignant growths. Chronic conditions such as CKD and acute kidney injury-AKI also represent significant kidney abnormalities that impact kidney function [2]. Each year, global kidney disease cases increase while nephrologists remain scarce. Kidney stone prevalence varies regionally, Asia has range in between 1 % and 19 %, 4 % in South America and the range in Europe varies between 5 % and 10 % [3]. Large-scale clinical screenings for kidney disease typically produce a vast number of medical images. Due to their time-intensive nature and the shortage of nephrologists, inaccurate diagnoses occur, leading to inadequate treatment for many patients [4]. To address this issue, computerized medical applications have been advanced to assist nephrologists in the early diagnosis of kidney diseases. Early detection often relies on various medical imaging techniques such as CT scans, ultrasound and MRI. Ultrasound imaging is particularly preferred for patient safety as it does not involve exposure to radiation [5].

Traditional methods for tumor prediction and diagnosis in the medical field often involve the use of Computer-Aided Diagnosis-CAD systems, which assist radiologists and other medical professionals in interpreting medical images [6]. These systems enhance the detection and characterization of tumors by highlighting suspicious areas in imaging modalities such as X-rays, MRI, and CT scans [7]. CAD systems typically employ image processing techniques to enhance the visibility of tumors, followed by feature [8] extraction and classification algorithms to distinguish between benign and malignant lesions [9]. For

E-mail address: saifurrehman.khan@csu.edu.cn.

https://doi.org/10.1016/j.compbiomed.2025.110214
Received 2 February 2025; Received in revised form 10 April 2025; Accepted 11 April 2025
Available online 14 April 2025

instance, in the evaluation of kidney stones, CAD systems analyze CT scans to identify and highlight potential calculi, assisting radiologists in detecting these formations which are indicative of kidney stones [10]. These systems enhance the precision of stone detection by distinguishing them from other dense structures within the kidney and urinary tract, thereby aiding in the accurate diagnosis and management of urolithiasis [11]. These systems do not replace the expertise of medical professionals but rather augment their capabilities by providing a second opinion and reducing the likelihood of oversight.

Kidney tumors can be classified into several types, primarily differentiated by their nature and origin. The most common type is renal cell carcinoma-RCC, which itself includes various subtypes such as clear cell RCC, papillary RCC, and chromophobe RCC [12,13]. Another significant category is urothelial carcinoma, which affects the renal pelvis [14]. For benign tumors, oncocytomas and angiomyolipoma are prevalent. The availability of diverse imaging modalities in datasets enhances the diagnosis and research into these tumors. Datasets often include modalities such as CT, MRI, and ultrasound, each providing unique insights into tumor characteristics.

The role of deep learning-DL in the diagnosis and management of kidney stones has become increasingly pivotal, leveraging its ability to automatically learn complex patterns from medical images [15]. DL models, particularly Convolutional Neural Networks-CNN, excel in identifying subtle features in imaging data that may indicate the presence of kidney stones, surpassing traditional image analysis techniques in both speed and accuracy [16]. The necessity for feature fusion arises from the multifaceted nature of medical images, where combining features from different layers of a DL model or different imaging modalities can significantly enhance diagnostic precision [17]. This fusion approach leverages the strengths of each feature set or modality, leading to a more comprehensive understanding of the data. The benefits of employing DL and feature fusion in the context of kidney stones include improved detection rates, enhanced ability to differentiate between stone types, and the potential for predicting treatment outcomes [18]. Applications extend beyond diagnosis to include monitoring the progression of the disease, guiding surgical planning, and customizing patient management strategies. This integration of DL technologies into the clinical workflow represents a transformative shift towards more accurate, efficient, and personalized care for patients with kidney stones.

In the domain of DL applied to medical imaging, both prediction fusion and feature fusion play critical roles in enhancing diagnostic accuracy, but they operate at different stages of the model architecture. Feature fusion involves combining various features or characteristics extracted from the data at an early or intermediate stage in the model [19]. This approach allows the model to leverage a richer set of information for making predictions, which is particularly beneficial when dealing with complex or subtle patterns in medical images, such as those found in kidney stone detection. On the other hand, prediction fusion occurs at a later stage, where predictions from multiple models or modalities are aggregated to form a final decision. This method is beneficial when individual models capture different aspects of the data, and their collective insights can lead to a more accurate [20]. The benefit of employing these fusion techniques lies in their ability to harness complementary information, either from within the same dataset or across different datasets and models, thereby improving the reliability and precision of medical diagnoses. Applications of these fusion strategies extend across various medical imaging tasks, including the detection and classification of kidney stones, where they contribute to more accurate and confident decision-making.

Main contribution of this work as follow:

- Study adopting feature fusion with an Inception block, the methodology reduces computational complexity while preserving rich feature representations. This enables the model to efficiently extract diverse spatial and temporal features, improving predictive performance without excessive overhead.
- The integration of ConvLSTM after feature fusion enables the model to capture temporal dependencies, improving sequential data analysis. The Inception block further enhances spatial feature extraction, leading to more robust hierarchical learning and improved classification accuracy.
- Study presents the combination of ConvLSTM and the Inception block further strengthens feature extraction by simultaneously leveraging temporal and spatial information. This approach improves hierarchical feature learning, allowing the model to effectively analyze patterns across different feature levels, leading to more robust and generalized predictions.
- A high-quality MHC-CT scan dataset is introduced, consisting of 1860 tumor and 1024 normal CT images, validated by medical experts. The dataset ensures high reliability and minimal misclassification risk, strengthening model generalizability and clinical applicability.

2. Related work

Recently, artificial intelligence-AI [21,22] has become a significant field in medical diagnostics, offering reductions in clinician workload and mitigating human errors. DL models are increasingly utilized in various medical domains such as skin [23], brain [24], retinal [25], and breast [26,27] disease detection. Similarly, DL techniques are applied in urology, particularly for automated kidney stone detection. However, there's a limited focus on leveraging these technologies for broader kidney disease diagnosis. Yildirim et al. [28] proposed an automated detection of kidney stones using a DL technique applied on coronal CT images. They used a total of 1799 CT images and achieved an accuracy of 96.82 % in detecting the presence of kidney stones, including small-sized stones from the CT images. K. Yildirim et al. [20] applied the Inception-V3 model, for kidney stone classification from CT scans, and achieving an accuracy of 98.52 %. However, the study faces a potential limitation due to the scarcity of authentic CT images available for training and testing. Aksakalli et al. [29] aimed to classify individuals using machine learning-ML and CNN. They tested different ML classifiers among which Decision Trees-DT achieved the highest F1-score as 85.3 % with S + U sampling, suggesting its feasibility for kidney x-ray image classification. However, limited by reliance on manual kidney stone detection.

Mahalakshmi et al. [30] proposed a kidney stone classification technique using optimized Transfer Learning-TL. Different DL models were utilized, with hyperparameters adjusted using Gorilla Troops Optimizer-GTO. The TL model achieved superior performance, with 98.49 % accuracy. Sudharson et al. [31] propose a CAD system using pre-trained ResNet-101 for feature extraction and SVM for classification. Achieving 87.31 % accuracy, the CAD system detects multi-class kidney abnormalities from ultrasound images. Despite superior existing methods, the CAD system reliance on pre-processing poses limitations. Dos Santos et al. [32] propose a DL model for kidney stone detection, achieving 96.20 % accuracy on computed tomography images. The model proves valuable by training on a dataset from urinary system examinations.

Chaki et al. [33] automate the detection of CT scans using an inductive TL ensemble DNN. They utilized three datasets for feature extraction from CT images of kidney stone by using pre-trained DNN models. The proposed strategy achieves 96.7 % accuracy using a dataset of noisy kidney CT images with binary classes normal and stone. Sabuncu et al. [34] utilized the Inception-V3 model pre-trained with other CNN architectures for abdominal CT scans of patients with kidney stones. Evaluating eight models with 8209 CT images, the Inception-V3 achieved 98.52 % test accuracy in detecting kidney stones but limited authentic CT images for training and testing pose a study limitation.

Yadav et al. [35] proposes a new model for precise prediction of CKD

Table 1
Comprehensive literature review highlighting key findings and limitations.

Reference	Method	Outcome	Limitation
[29]	Decision Trees	Achieved the highest F1-score as 85.3 % with S + U sampling	Limited by reliance on manual kidney stone detection
[30]	Optimized DL model with GTO	Optimized TL models, achieving a high accuracy of 98.49 %	GTO presents slow convergence in high-dimensional search spaces
[31]	ResNet-101 with SVM	CAD system detects multi-class kidney abnormalities with 87.31 % accuracy	Despite superior existing methods, the CAD system reliance on pre-processing poses limitations
[32]	DL	The proposed model attained an accuracy of 96.20 % on the binary dataset	Restricted model generalization assessment
[34]	Inception-V3	Achieved 98.52 % test accuracy in detecting kidney stones	The scarcity of authentic CT images for training and testing presents a limitation in this study
[35]	Hybrid flash butterfly optimization algorithm	The proposed model achieves exceptional performance, attaining an accuracy of 97.8 %	HFBOA may introduce instability in convergence. If the algorithm overly emphasizes exploration (global search), it may struggle to converge to an optimal solution efficiently.

which is evaluated on CKD dataset. They select the features by using a hybrid flash butterfly optimization algorithm-HFBOA and optimize the weight functions. The proposed model demonstrates superior performance with an accuracy of 97.8 %. Wu et al. [36] present an automated architecture for detecting kidney abnormalities from abdominal ultrasound images using convolutional neural networks. A dataset of 3722 abdominal ultrasound images with annotations is utilized for training and evaluation, achieving an average classification accuracy of 94.67 % with the Mf-Net model. Rao et al. [37] present a fusion DL model that combines a Graph Neural Network-GNN with a structural data to predict CKD progression with 95.08 % accuracy, highlighting the benefits of integrating graph-structured data. The CKD dataset used in this study is sourced from the India UCI ML Repository, comprising 400 samples, each containing 24 features and a binary class label. Table 1 provides a comprehensive literature review, summarizing key findings and limitations.

3. Methodology

Larger datasets tend to yield superior results for conventional CNN base approaches compared to smaller ones. However, in scenarios where the dataset is limited in size, TL emerges as a valuable strategy. TL [23, 38] involves leveraging a pre-trained model from a larger dataset and applying it to a smaller dataset. This approach eliminates the necessity for a substantial dataset and significantly reduces the time-consuming training process required when building a DL model from the scratch.

3.1. Dataset collection

In this study, we have curated an exclusive dataset termed the Multiple Hospital Collected MHC-CT scan, consisting of tumor and non-tumor samples. The dataset comprises a binary classification task, with 1860 tumor CT scan images and 1024 normal CT scan images, each class exhibiting standard CT-scan image views. Every image has undergone meticulous labeling by medical professionals, with validation conducted by multiple doctors to minimize the possibility of misclassification. This dataset was ethically assembled from 120 patients over a span of six

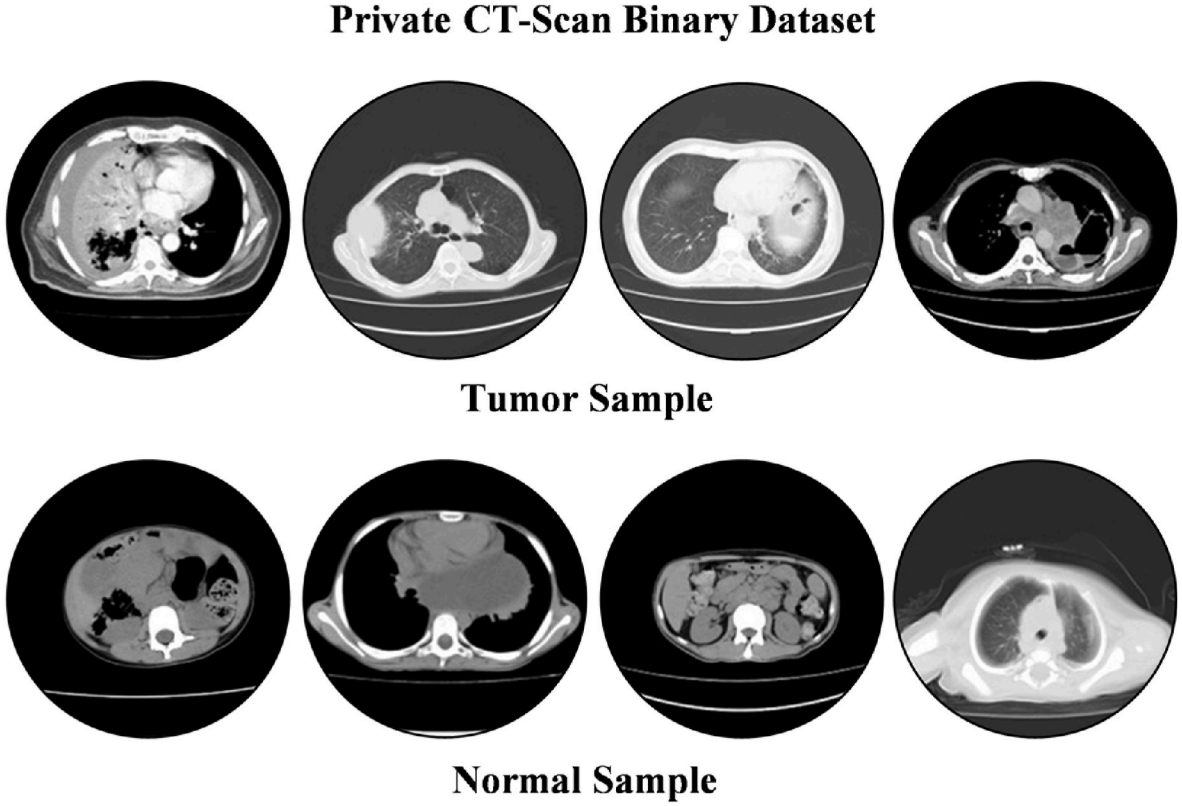


Fig. 1. Overview of MHC-CT scan image (private): Binary Class.

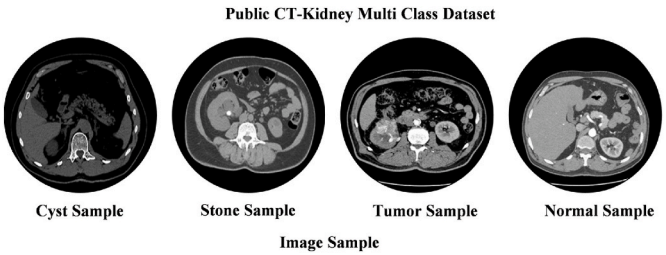


Fig. 2. Overview of publicly available dataset: Multi Class.

Table 2
Overview of augmentation steps employee in this study.

Augmentation	Details	Benefits
Rescale	Each image in our dataset underwent rescaling, accomplished by dividing the pixel values by 255.	This normalization process ensures that the images fall within the standard dimension range of [0,1], thereby preparing them for effective processing by the DL model.
Flipping	Applied both vertical and horizontal flipping on the training images.	A horizontal flip creates mirrored versions of each original image by enabling the model to identify objects from both left-to-right and right-to-left orientations. A vertical flip generates top-to-bottom mirrored versions of each image, aiding the model in learning object recognition from vertical reversals as well.
Rotation	Applied a 90-degree rotation to each image in both datasets.	This method creates more discrepancies of the original images. Implementing rotations enhances the ability of model to recognize objects from various orientations, thereby increasing its robustness.

months, with an additional two months dedicated to verifying the labeling accuracy. The utilization of this newly created dataset presents several advantages over existing datasets. Firstly, the comprehensive labeling process involving multiple medical professionals ensures a higher degree of accuracy and reliability in classification. Secondly, the inclusion of a diverse range of patient samples from multiple hospitals enhances the dataset representativeness and generalizability, thus development more robust and applicable research outcomes in the field of CT scan analysis. Fig. 1 displays sample images from the MHC-CT scan dataset.

To further boost the assessment of our proposed model robustness, we have utilized a widely recognized benchmark multi-class CT scan dataset [28] that's publicly available. Validating our model performance with both public and private datasets enhance its reliability and generalizability. Public datasets offer transparency and reproducibility, while private datasets provide a real-world test, ensuring effectiveness across diverse data distributions.

We used the CT KIDNEY DATASET [39] which was collected through Picture Archiving and Communication System-PACS from different hospitals in Dhaka, Bangladesh. The dataset includes patients diagnosed with kidney tumors, cysts, stones, or normal conditions. It contained total 12,446 images including 3709 for cysts, 5077 for normal cases, 1377 for stones, and 2283 for tumors. The data is divided into four categories: Normal, Cyst, Tumor, and Stone. Fig. 2 displays sample images from the publicly available multiclass dataset.

3.2. Preprocessing and augmentation

The CT image dataset for kidney stones underwent preprocessing to improve the accuracy of the proposed architecture, as it often contains background noise. Normalizing the CT images to a standard size is important because different DL pre-trained models have specific data format requirements. For example, the VGG19 or ResNet50 framework necessitates images to be of 224 × 224 pixels. Consequently, every kidney stone image has adjusted to this specific size to comply with the input necessities of the pre-trained DL models, which is typically

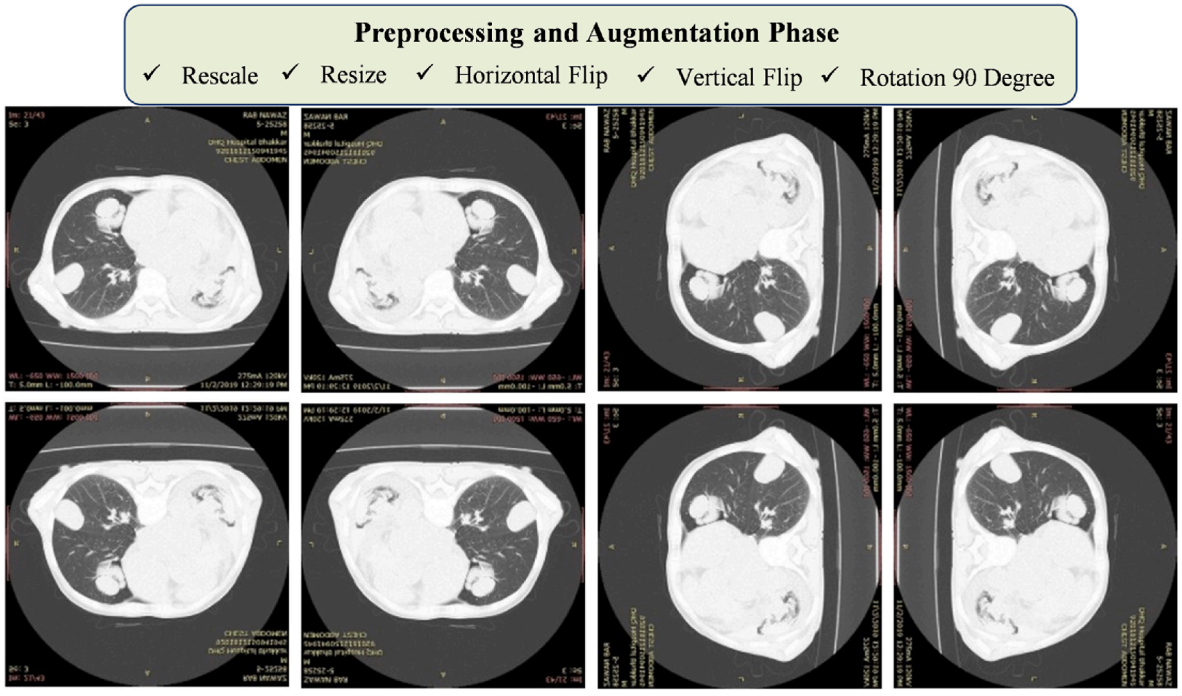


Fig. 3. Overview of augmentation steps: Private- Binary Class.

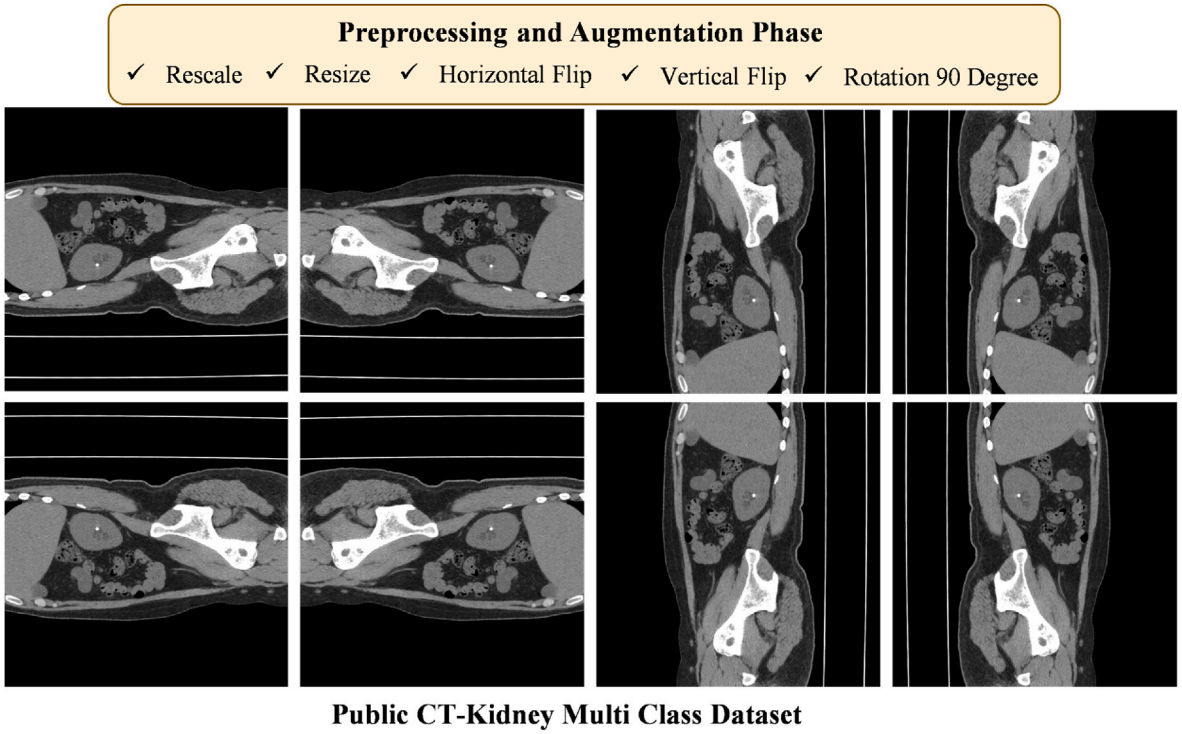


Fig. 4. Overview of augmentation steps: Public Multi Class.

Table 3
Dataset distribution across both dataset.

Dataset	Category	Train	Validate	Test	Total
Private (MHC-CT)	Binary	2314	286	286	2886
Public (CT KIDNEY DATASET [39])	Multi	9960	1243	1243	12446

preferred as square-shaped images. This standardize has improved computational efficiency, supported broader generality, and simplified the integration of pre-trained models, which will enhance the performance of the models.

In this research, we utilized image augmentation strategies to address the issue of an imbalance dataset and prevent the risk of overfitting. Instead of gathering additional data, our focus on enriching the current dataset, thereby offering pre-trained models a broader and more varied set of training images. Our objective to enhance the classification performance of the DL model. Table 2. Presents the detail of each augmentation step. Figs. 3 and 4 display the augmented sample images from both datasets: the Private Binary Class and the Public Multi Class. Table 3 displays the distribution of the dataset across three groups: Train (80 %), Validation (10 %), and Test (%10).

Using image augmentation techniques like rescaling, rotating, and flipping increases the diversity of the training data which is crucial for enhancing the robustness of a DL model. These techniques prevent overfitting by training the model on varied representations of the same images, thus improving its ability to generalize to a new unseen data. Additionally, augmentations such as rotations and flips enable the model to recognize objects in images regardless of their orientation or position, which significantly enhancing its classification performance.

3.3. Feature extraction with CNN based model

In medical imaging, classifying and analyzing images are critical but challenging tasks. These challenges arise mainly because there are not many extensive annotated images datasets and there's also a high variability in medical images. Pre-trained models emerge as a vital solution in this context because they can leverage data and features learned from extensive and diverse datasets that were not specific to medical applications. These models often start with training on large datasets like ImageNet and are then adapted for medical imaging tasks through techniques such as TL, which adapts them to handle medical data effectively. This TL approach allows for more accurate and faster results even with smaller datasets typical in medical settings. Pre-trained models also reduce the need for extensive computational resources that are often necessary when training models from scratch. By leveraging these efficient models, medical professionals can achieve higher diagnostic accuracy, which is crucial for effective patient care. Pre-trained models offer significant advantages in the classification of medical images compared to traditional models.

Using pre-trained models helps with regularization and generalization by providing a robust starting point that prevents overfitting on smaller, specific datasets. This approach extends the model ability to perform well across various unseen medical data, enhancing its general applicability. The benefits of using pre-trained models in medical image classification are substantial. Firstly, they significantly reduce the need for large medical-specific datasets, which are often unavailable or costly to obtain. By using models that have already learned complex patterns from vast amounts of general images, medical imaging tasks can be approached with a head start. Secondly, pre-trained models can accelerate the development process, enabling faster deployment of medical imaging solutions. This speed is crucial in medical diagnostics, where timely decision-making can dramatically impact patient outcomes. Lastly, these models enhance the accuracy of classifications. Pre-trained models bring a depth of learned features that can be fine-tuned to detect subtle nuances in medical images, leading to more reliable and precise diagnostic outcomes.

3.4. ResNet50 model

The ResNet50 architecture is a type of CNN known for its depth and efficiency, introduced by Kaiming He and colleagues in 2016 on deep

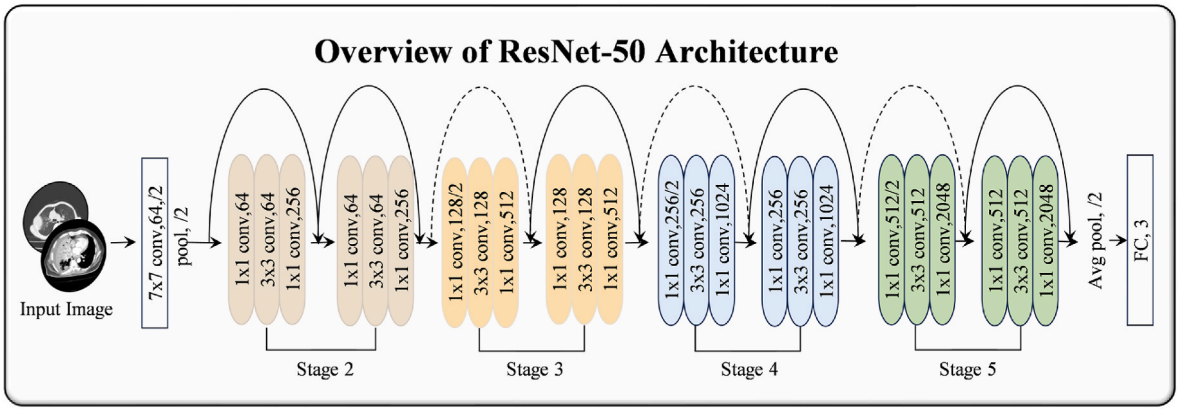


Fig. 5. Illustrate the architecture overview of the ResNet-50-base model.

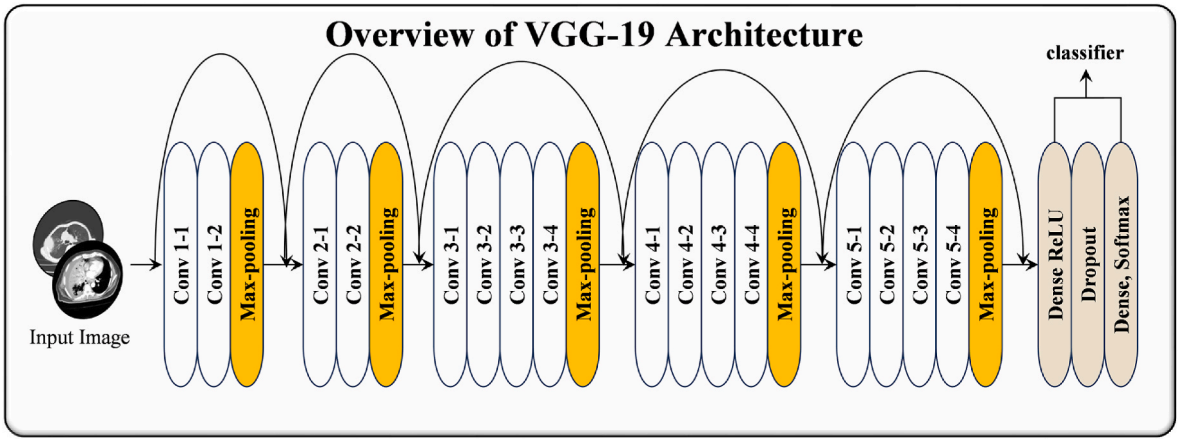


Fig. 6. Illustrate the architecture overview of the VGG-19-base model.

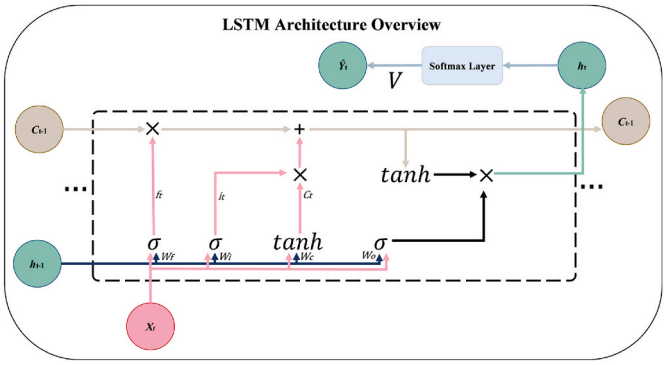


Fig. 7. Illustrate the architecture overview of the ConvLSTM architecture.

residual networks [40]. This model belongs to the ResNet series, which initiated the concept of RL to ease the training of networks that are substantially deeper than those used previously. ResNet50 consists of 50 layers, including convolutional layers, pooling layers, and a fully connected layer at the end. The core idea behind ResNet50 is the use of residual blocks. These blocks contain skip connections that allow the input to a layer to be added to its output. This setup helps in combating the vanishing gradient problem by allowing gradients to flow through the network directly. Each residual block in ResNet50 typically consists of three layers. The first and third layers are 1×1 convolutions, responsible for reducing and then increasing dimensions, thereby preserving the network's depth. The second layer uses 3×3 convolutions. This approach minimizes the parameter count and computational complexity. The network uses batch normalization and ReLU activation functions after every convolution operation except for the final output from the residual blocks where the addition with the shortcut connection is performed first before applying ReLU.

The presence of these skip connections allows ResNet50 to learn identity functions that ensure that higher layer outputs are at least as good as lower layer outputs. This architecture has shown significant improvements in terms of speed and accuracy in various image recognition tasks and has been a pivotal model in advancing DL applications. Fig. 5 illustrate the architecture overview of the ResNet-50-base model.

3.5. VGG19 model

The VGG19 architecture detailed in the foundational paper by Karen Simonyan and Andrew Zisserman is a deep convolutional neural network with 19 layers [41]. The VGG19 network includes 16 convolutional layers and 3 fully connected layers [42]. The convolutional layers in the VGG19 architecture use 3×3 filters, which is the smallest size needed to capture the notion of left/right, up/down, and center directions. The stride for these convolutions is fixed at 1 pixel, and the layers are padded by 1 pixel to maintain the spatial resolution after convolution. The architecture includes several max pooling layers which are placed specifically after certain convolutional layers to reduce the spatial size of the representation. This reduction helps decrease the number of parameters and the computational load of the network. The fully connected layers follow the stack of convolutional layers, and the final layer is a SoftMax classification layer that outputs the probabilities of the 1000 class labels. The network employs the ReLU activation

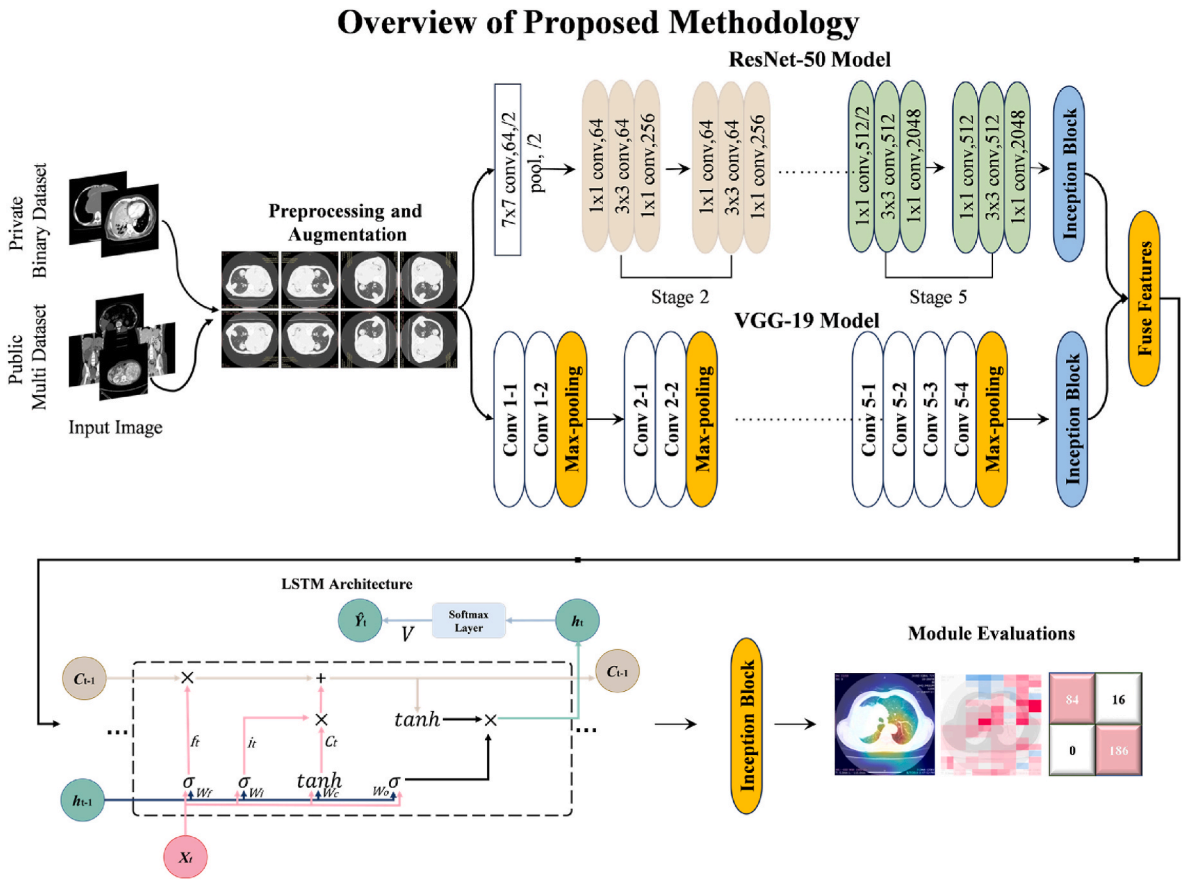


Fig. 8. Overview of the proposed methodology framework.

Table 4
Hyperparameter with details.

Parameters	Details
Batch-Size	64
Learning-Rate	1-e^{-3}
Epochs	20
Optimizer	Adam

Table 5
Class wise performance evaluation of proposed model on private dataset: Binary Class.

Models	Class	Precision	Recall	F1-Score	Accuracy
VGG19-Base	Normal	0.95	0.99	0.97	97.90 %
	Tumor	0.99	0.97	0.98	
ResNet50-Base	Normal	0.88	0.98	0.92	94.40 %
	Tumor	0.99	0.92	0.96	
Fusion	Normal	0.95	1.00	0.98	98.25 %
	Tumor	1.00	0.97	0.99	
ConvLSTM with Inception block	Normal	0.99	0.98	0.98	98.60 %
	Tumor	0.97	1.00	0.99	
Proposed Model	Normal	0.98	1.00	0.99	99.30 %
	Tumor	1.00	0.99	0.99	

function to introduce non-linearity, enabling it to learn more complex patterns in the data. This architecture is designed to capture a wide range of features at multiple levels, making it highly effective for large-scale image recognition tasks. Its capabilities are clearly demonstrated by its strong performance in the ImageNet challenge. Fig. 6 illustrate the architecture overview of the VGG19-base model.

3.6. Optimized model interpretation with feature fusion

In this study, we aimed to enhance hierarchical feature extraction efficiency while preserving crucial features by incorporating pre-trained model techniques into our framework. Departing from conventional methods, we systematically assessed various pre-trained models such as DenseNet, Xception, and MobileNet, ultimately selecting ResNet50 and VGG19 for their high performance in initial setup. To augment deep feature extraction, we employee an inception block, facilitating the simplification of complex features into more manageable components. The integration of inception block into each model before feature fusion presents an innovative approach to addressing the limitations of direct fusion prediction. This block is designed to capture multi-scale features, thereby enhancing the model ability to discern intricate patterns and structures within the data. An in-depth analysis of the computational overhead associated with incorporating this block and impact on overall model performance would offer valuable insights into its efficacy. Disadvantages of direct fusion prediction with pre-trained models may include potential loss of nuanced information due to rigid integration strategies, hindering the model adaptability to diverse datasets and scenarios. So, this block aid in capturing multi-scale features, enhancing the model ability to discern intricate patterns and structures within the data. Additionally, enhance model in feature extraction, allowing for improved generalization across varying input conditions. Fig. 8 illustrate the feature fusion along inception block approach.

3.7. ConvLSTM layer

In this study, Integrating a ConvLSTM layer [43] after feature fusion along an inception block offers several benefits. Such as, ConvLSTM units excel in capturing long-term dependencies in sequential data,

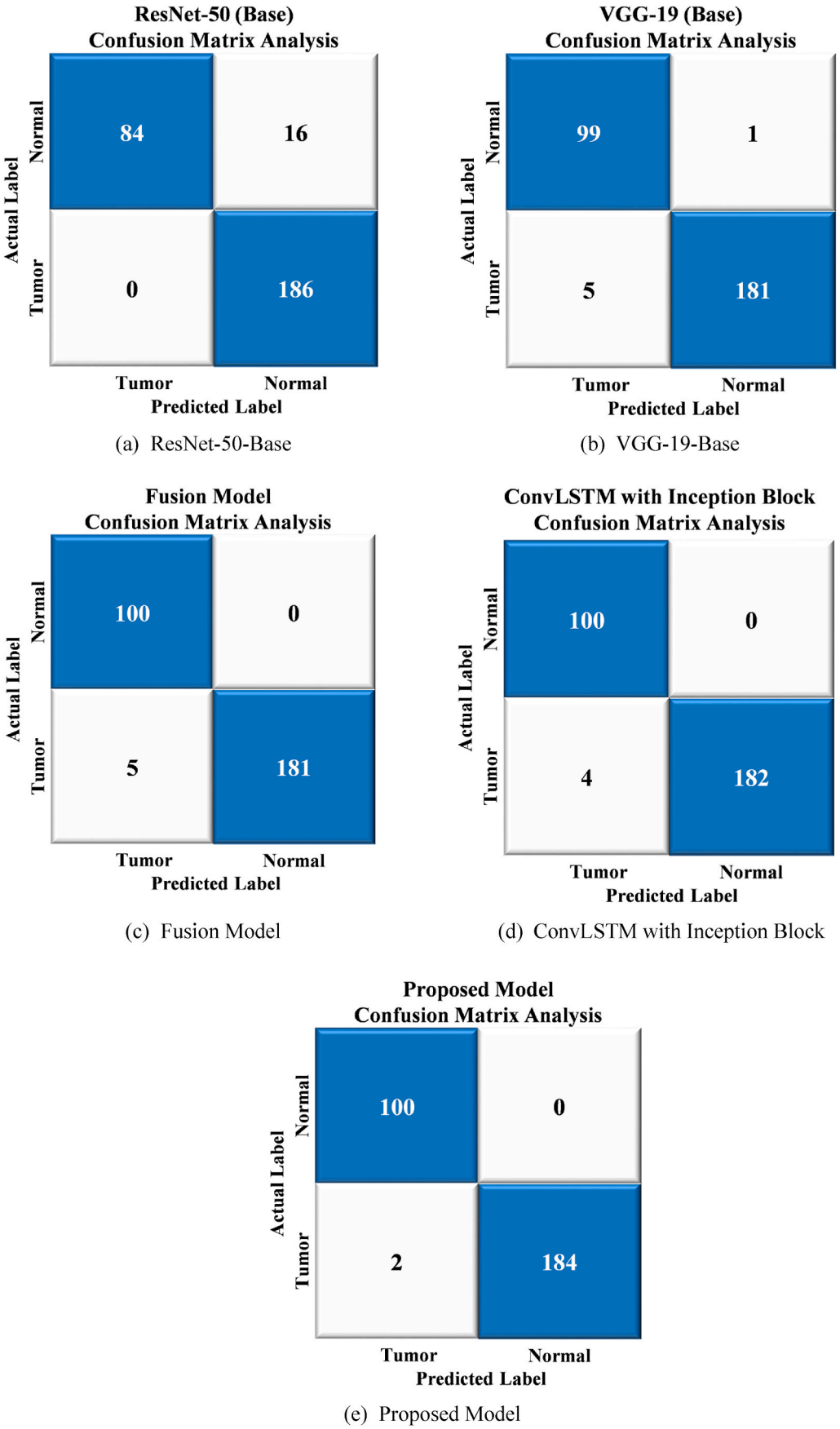


Fig. 9. Confusion matric visualization using private: Binary Class: (a) ResNet-50-Base, (b) VGG-19-Base, (c) Fusion model, (d) ConvLSTM with Inception Block and (e) Proposed Model.

Table 6
Class wise performance evaluation of proposed model on public dataset: Multi Class.

Models	Class	Precision	Recall	F1-Score	Accuracy
VGG19-Base	Cyst	0.94	0.84	0.89	86.96 %
	Normal	0.87	0.99	0.92	
	Stone	0.60	0.64	0.62	
	Tumor	0.97	0.79	0.87	
ResNet50-Base	Cyst	0.91	0.90	0.90	86.88 %
	Normal	0.89	0.93	0.91	
	Stone	0.68	0.55	0.60	
	Tumor	0.85	0.87	0.86	
Fusion	Cyst	0.92	0.85	0.88	87.77 %
	Normal	0.90	0.96	0.93	
	Stone	0.66	0.65	0.66	
	Tumor	0.89	0.86	0.88	
ConvLSTM with Inception Block	Cyst	0.91	0.86	0.93	89.30 %
	Normal	0.89	0.99	0.90	
	Stone	0.66	0.83	0.72	
	Tumor	0.99	0.75	0.83	
Proposed Model	Cyst	0.99	0.86	0.93	91.31 %
	Normal	0.89	1.00	0.96	
	Stone	0.69	0.88	0.76	
	Tumor	1.00	0.79	0.89	

enhancing the model's ability to understand temporal relationships in the fused features. Moreover, by incorporating ConvLSTM layer, the model gains the capability to selectively retain or discard relevant information from the fused features, improving the overall robustness and interpretability of the model. Addition to it, ConvLSTM inherent ability to mitigate the vanishing gradient problem ensures more stable and efficient training, leading to better convergence and higher performance of the model in classification tasks such as feature sequence prediction. The basic architecture diagram (Fig. 7) of a ConvLSTM unit typically consists of four main components: the input-gate, the forget-gate, the output-gate, and the cell-state. These components work together to regulate the flow of information through the unit. The input gate controls the information that enters the cell-state, the forget-gate determines which information to discard from the cell-state, and the output-gate regulates the information that gets passed to the output. The cell-state serves as the memory of the unit, allowing it to retain information over time and mitigate the vanishing gradient problem often encountered in traditional recurrent neural networks.

3.8. Overview of proposed model

In the medical field, accurate diagnosis is paramount for early detection of diseases. Computer vision, particularly through analyzing CT scan images, plays a crucial role in this process. In our study, we introduce a novel methodology for predicting tumor diseases. Initially, we evaluated six pretrained models under the same experimental conditions to ensure fair comparison. Among these models, ResNet50 and VGG19 emerged as the top performers. ResNet50 excels due to its deep residual learning, enabling improved training and convergence, while VGG19 deeper architecture enhances feature extraction and representation capabilities. Unlike many studies that rely on multiple model fusion for prediction, we avoid potential computational complexities and information loss by adopting feature fusion along with an inception block. Inception block efficiently extract diverse features by combining various filter sizes and pooling operations within a single layer, enhancing feature representation. This approach facilitates the integration of multi-scale information while maintaining imbalance dataset efficiently. Furthermore, we enhance the model robustness and accuracy by incorporating ConvLSTM unit after feature fusion with an inception block. ConvLSTM unit excel in capturing long-term dependencies in sequential data, making them ideal for tasks involving sequential feature prediction, and other sequential data analysis. By leveraging ConvLSTM unit in conjunction with feature fusion, our proposed model can capture nuanced patterns and context, leading to more accurate predictions. By integrating these techniques such as feature fusion with inception blocks and ConvLSTM units, our model not only achieves high accuracy in kidney disease prediction but also offers robustness and generalizability. This enables clinicians to make more informed decisions and potentially detect kidney disease at earlier stages, improving patient outcomes and treatment efficacy.

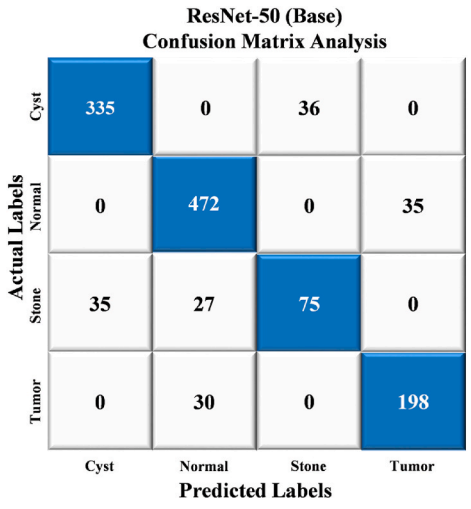
In addition to its high accuracy and robustness, our proposed model offers several benefits to conventional models. Its streamlined approach of feature fusion along with inception blocks minimizes computational complexity while maximizing information extraction. This efficiency translates to faster processing times, making the model suitable for real-time applications in clinical settings. Moreover, the model ability to capture long-term dependencies through ConvLSTM unit enhances its predictive capabilities, particularly in scenarios where sequential data play a crucial role, such as tracking disease progression over time or analyzing patient histories. Furthermore, the model interpretability is improved through its utilization of well-established neural network architectures like ResNet50 and VGG19, making it easier for clinicians to understand and trust the predictions provided. Fig. 8 illustrate the overview of proposed methodology.

4. Implementation and results discussion

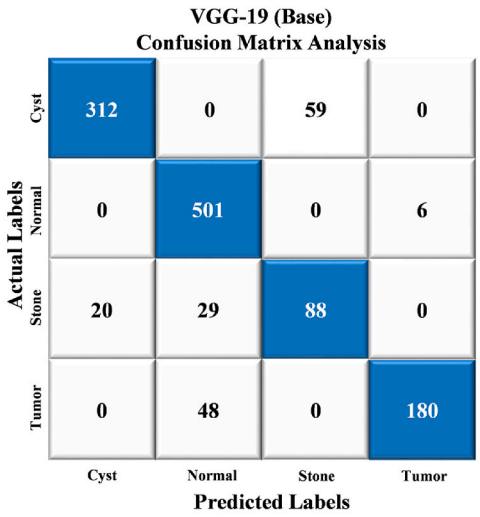
In this section, we delve into our model implementation, highlighting its intricacies. We then present experimental findings that underscore its performance, conducting a thorough GRAD-CAM analysis, Shap visualization, and Feature Map examination to provide insight into its decision-making process. Finally, we compare its outcomes with those of previous methodologies, pre-trained models, and proposed models across various configurations.

4.1. Model hyperparameter detail

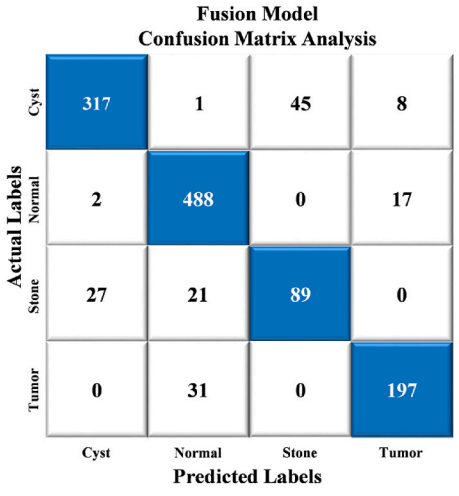
In training DL models like ResNet-50 and VGG-19, selecting the right hyperparameters is crucial for optimizing performance. Hyperparameters are adjustable parameters that must be set before the training process begins and they directly influence how training is conducted. The choice of these parameters can significantly affect the speed of learning process. Batch size is one such hyperparameter that determines the quantity of training samples processed prior to updating the model internal parameters. The need for an optimal batch size arises from the trade-off between memory usage and the stability of the learning process. A larger batch size ensures more stable gradient estimates, but it requires more memory and can make the training process computationally expensive. The learning rate is a crucial hyperparameter that controls the extent of adjustments made to the model based on the estimated error during each update of the model weights. Selecting an appropriate learning rate is essential, as a rate that is too low may result in an excessively prolonged training period with the potential of becoming stuck, while a value too high can cause the training to oscillate around or diverge from the optimal set of weights. Epochs refer to the number of times the learning algorithm will work through the entire training dataset. The number of epochs is important because too few epochs can result in an underfit model, whereas too many epochs can lead to overfitting. Therefore, setting the right number of epochs is crucial for balancing training performance and model accuracy. Optimizers are techniques or strategies employed to adjust the neural network's parameters such as weights and learning rates in order to minimize losses. Table 4 presents the details of hyperparameter utilized in this study. After analyzing the initial experimental results, we have refined our hyperparameter selection to optimize performance. To select the optimal model hyperparameters, we systematically incorporated all hyperparameter settings highlighted in the literature and evaluated them individually. After testing each configuration, we



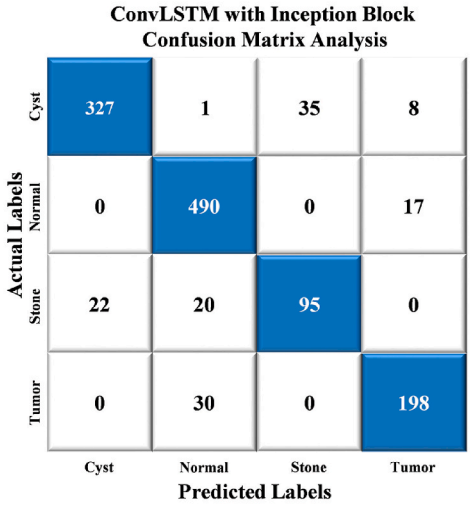
(a) ResNet-50-Base



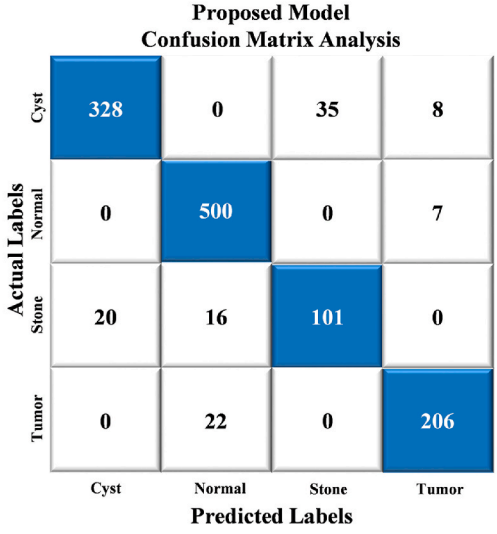
(b) VGG-19-Base



(c) Fusion Model



(d) ConvLSTM with Inception Block



(e) Proposed Model

Fig. 10. Confusion matric visualization using public: Multi Class: (a) ResNet-50-Base, (b) VGG-19-Base, (c) Fusion model, (d) ConvLSTM with Inception Block, and (e) Proposed Model.

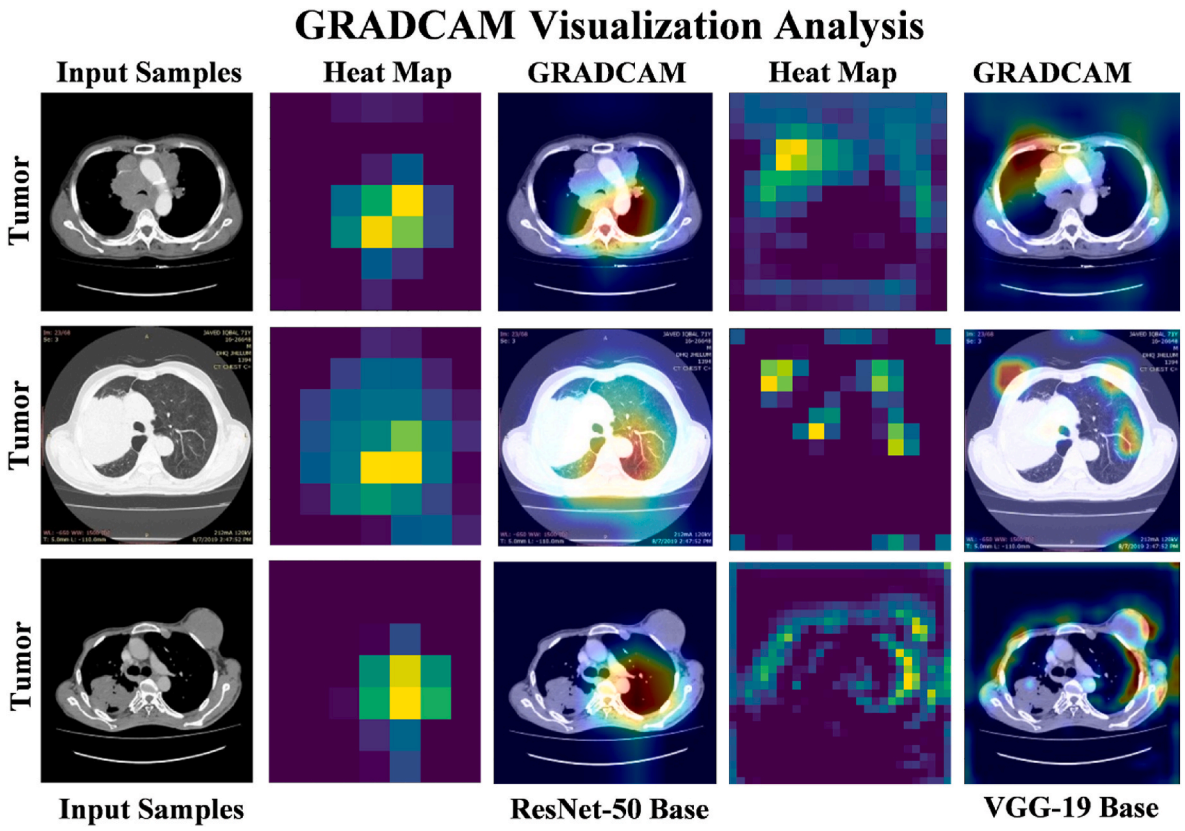


Fig. 11. Overview of GRAD-CAM visualization: Private- Binary Class.

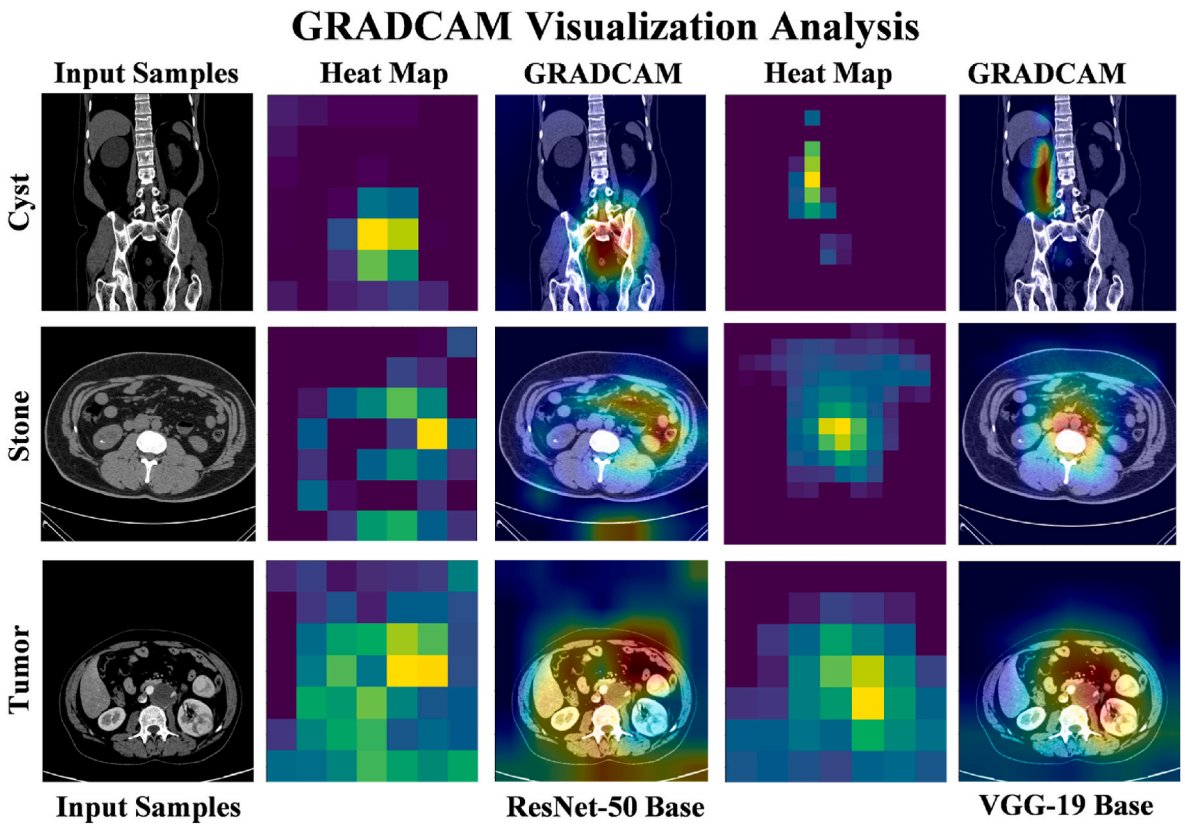


Fig. 12. Overview of GRAD-CAM visualization: Public- Multi Class.

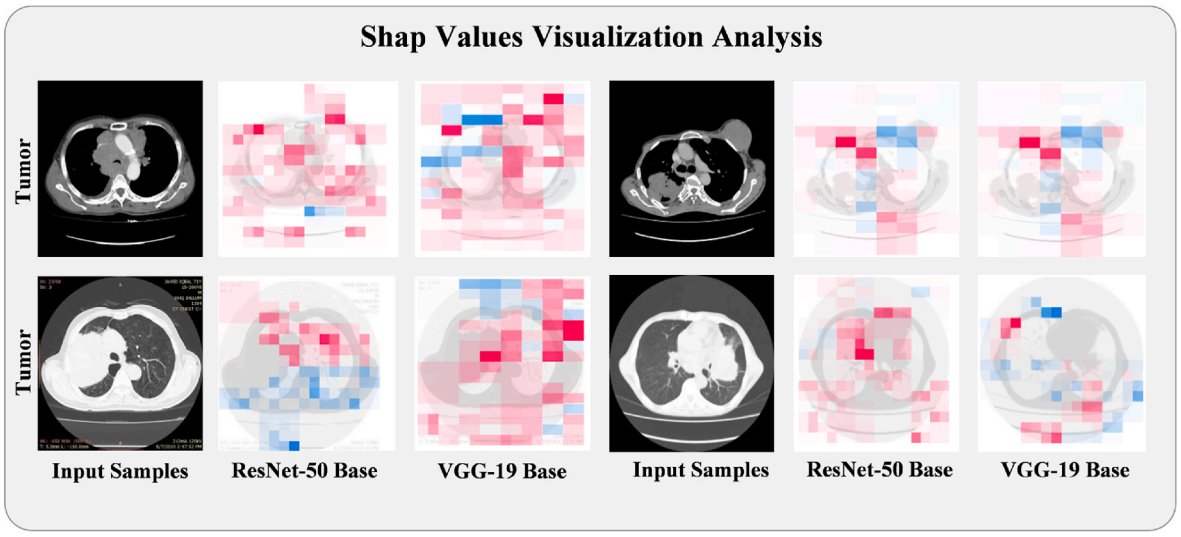


Fig. 13. Overview of Shap visualization: Private- Binary Class.

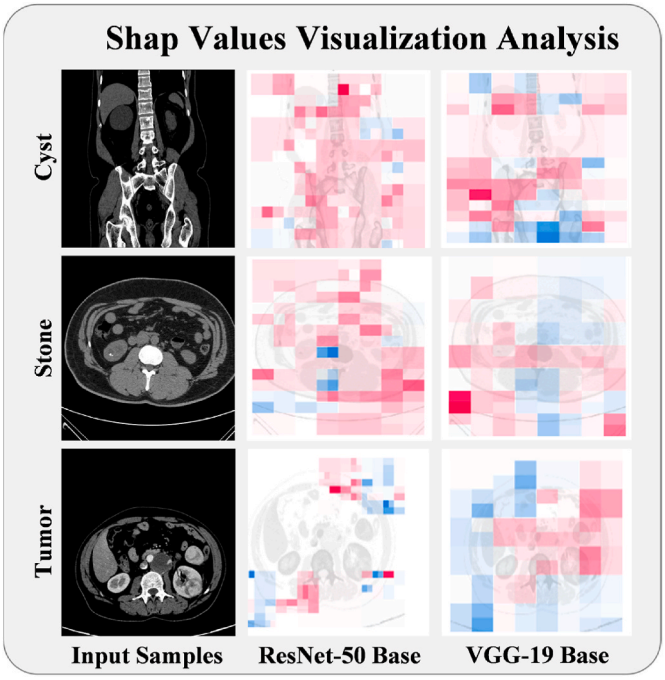


Fig. 14. Overview of Shap visualization: Public- Multi Class.

adopted the hyperparameter setting that achieved the highest model performance across all those mentioned in the literature review.

4.2. Model evaluation metrics

Model evaluation metrics are essential tools used to calculate the performance of DL models. The performance of the proposed model has assessed through metrics based on the confusion matrix. Accuracy in measures the proportion of correct predictions among the total number of cases evaluated. It is an essential metric for overall model performance but may not always offer a complete representation, especially in imbalanced datasets (eq-1). Precision focuses on the proportion of true positive predictions in the positive class, making it crucial for scenarios where the cost of false positives is high (eq-2). Recall measures the ability of the model to capture all relevant cases by calculating the proportion of true positives identified from all actual positives. This metric is vital in situations where missing a positive case has significant consequences (eq-3). The F1 score harmonizes precision and recall into a single metric by taking their harmonic mean, offering a balance between the two and providing a more comprehensive evaluation of model performance, especially in cases of uneven class distribution (eq-4). Together these metrics provide a multifaceted view of a model effectiveness and guide improvements and adjustments.

$$Accuracy : \frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

$$Precision : \frac{TP}{TP + FP} \tag{2}$$

$$Recall : \frac{TP}{TP + FN} \tag{3}$$

$$F1 - Score : 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

4.3. Model performance evaluation with private dataset: binary class

In the current study, the proposed model emerges as a novel solution for diagnosing kidney stone patients using CT images with minimal radiologist involvement. This model marks a significant advancement in efficiency, and reduce the manual tests time. In the pursuit of advancing medical image classification, our research has meticulously evaluated the performance of various DL models.

Table 5 showcases the class wise performance evaluation of proposed model on a private dataset categorized into binary classes such as Normal and Tumor. The VGG19 base model achieves precision scores of 0.95 for normal and 0.99 for tumor. Its recall scores are equally high at 0.99 for normal and 0.97 for tumor, leading to F1-scores of 0.97 and 0.98 respectively. Overall, VGG19 attains an accuracy of 97.90 %. In comparison, the ResNet50 base model records a lower precision of 0.88 for Normal while maintaining 0.99 for tumor. The recall rates are 0.98 for normal and 0.92 for tumor, resulting in F1-scores of 0.92 and 0.96. ResNet50 achieves an overall accuracy of 94.40 %.

The Fusion model scores a precision of 0.95 for normal and a perfect 1.00 for tumor. The model also excels in recall, achieving 1.00 for normal and 0.97 for tumor, which results in F1-scores of 0.98 and 0.99 respectively. It achieves an accuracy of 98.25 %. For a ConvLSTM with inception block. It achieves high performance than Fusion, particularly notable in detecting tumor with precision of 97 %, recall of 100 %, and

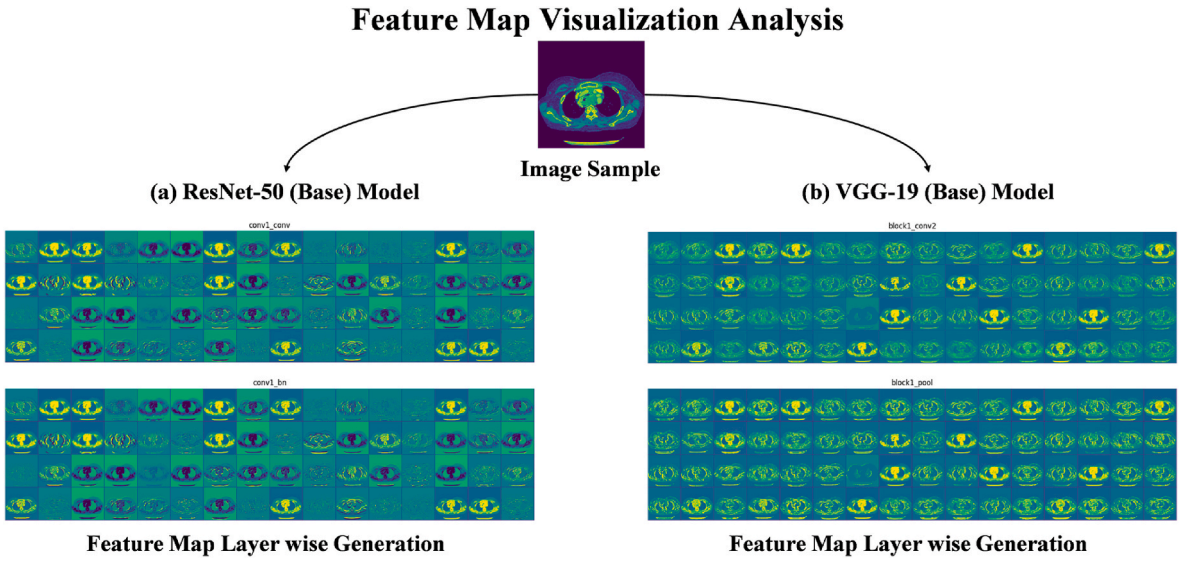


Fig. 15. Overview of feature map visualization: Private- Binary Class.

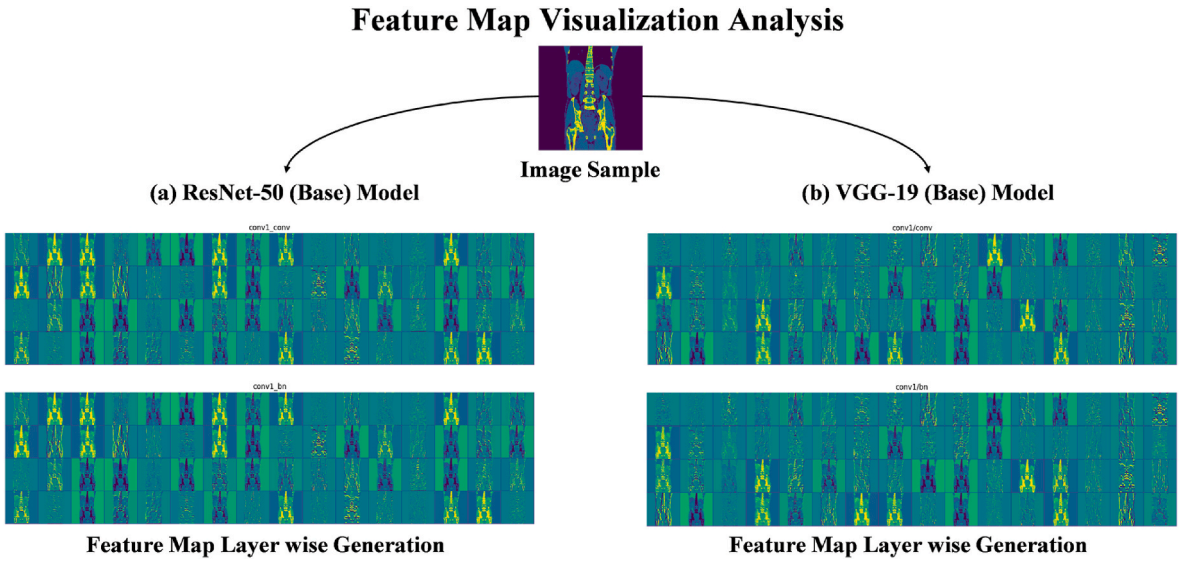


Fig. 16. Overview of feature map visualization: Public- Multi Class.

Table 7
Performance evaluation of proposed model on private dataset with pre-trained model: Binary Class.

Models	Precision	Recall	F1-Score	Accuracy
DenNet201 (Base)	95.81	91.50	93.17	94.05
ResNet101 (Base)	89.48	92.08	90.36	90.90
ResNet152 (Base)	92.00	91.81	91.90	92.65
Xception (Base)	95.03	93.60	94.30	94.08
Proposed Model	99.01	99.46	99.23	99.30

Table 8
Performance evaluation of proposed model on public dataset with pre-trained model: Multi Class.

Models	Precision	Recall	F1-Score	Accuracy
DenNet201 (Base)	78.63	78.14	77.72	81.49
ResNet101 (Base)	78.12	76.40	77.10	83.58
ResNet152 (Base)	83.17	77.90	79.48	85.11
Xception (Base)	80.59	78.88	79.41	82.46
Proposed Model	90.48	91.02	92.98	91.31

F1-score of 99 %. Overall accuracy is 98.60 %. The proposed model outperforms all with a precision of 0.98 for normal and 1.00 for tumor. It also excels top recall scores of 1.00 for normal and 0.99 for tumor, resulting in a consistent F1-score of 0.99 for both classes. It achieves an overall accuracy of 99.30 % among other base models. These results demonstrate the proposed model superior accuracy and reliability in classifying normal and tumor classes compared to the baseline models. This performance not only highlights the effectiveness of the proposed model but also underscores the potential for DL to develop medical diagnostics, providing tools that are both highly accurate and reliable for clinical decision-making.

Fig. 9 illustrates the confusion matrix visualization of the proposed model compared to backbone models, Fusion, ConvLSTM with inception block and proposed model. The visualization distinguishes between true and false labels. Fig. 9(a) illustrates the ResNet-50 (Base) confusion matrix, reporting 84 true negatives, 186 true positives, 16 false

Table 9
Performance comparison with existing studies.

Reference	Models	Binary Class	Multi class
Hossain et al. [48]	ResNet50/Kidney Disease Dataset	–	87.90 %
Srivastava et al. [49]	weighted average ensemble model	98.75 %	–
Majid et al. [50]	Fine-tuned DL model	94.09 %	–
Özbay et al. [51]	SSLSD-KTD method	98.04 %	
Proposed (Our)	Proposed Model	99.30 %	91.31 %

positives, and 0 false negatives (FN). Fig. 9(b) presents the VGG-19 (Base) confusion matrix, showing that the model correctly classifies 280 out of 286 test samples. Fig. 9(c) illustrates the fusion model performance, outperforming individual models by correctly classifying all normal samples while misclassifying only five tumor samples. On the other hand, Fig. 9(e) the proposed model outperforms all aforementioned models, with only two misclassifications in the tumor class, totaling 186 instances.

4.4. Model performance evaluation with benchmark dataset: Public Multi Class

Our research has thoroughly evaluated the performance of various DL models to advance medical image classification. We have already assessed the proposed model on a binary classification dataset. Now, we will examine its effectiveness on a multiclass classification using a benchmark dataset. Table 6 demonstrate the class-wise performance assessment of base models and proposed model on benchmark multiclass dataset having four classes including Cyst, Normal, Stone, and Tumor. This evaluation will help determine how well the model performs across a broader range of classes. Section 4.2 provides the evaluation metrics of each model for every class. Precision metric indicates the accuracy of positive predictions. A higher precision means a model returns more relevant results. Recall measures a model capability to identify all relevant instances within a dataset. Higher recall means fewer true positives are missed. The F1-Score represents the harmonic mean of precision and recall, striking a balance between the two. It especially helpful when dealing with imbalanced class distributions. Accuracy is the overall percentage of correctly predicted instances out of all predictions made. The models evaluated include VGG19 (Base), ResNet50 (Base), Fusion, and a Proposed Model.

The VGG19-Base model demonstrates varying effectiveness for different classes. For the Cyst class, its accuracy and reliability in identifying relevant cases result in a score of 0.89, while it scores 0.92 in accurately and reliably identifying Normal cases. Its effectiveness drops for the Stone class with a lower score of 0.62. However, it shows strong results in the Tumor class with a score of 0.87. Overall, this model achieves a comprehensive accuracy rate of 86.96 %. The ResNet50-Base model also shows varied capabilities. In the Cyst class, it effectively and reliably identifies relevant cases with a score of 0.90. It maintains high effectiveness in the Normal class with a score of 0.91 but shows less impressive results in the Stone category with a score of 0.60, and the Tumor category with a score of 0.86. The overall accuracy rate for this model is slightly lower at 86.88 %.

The Fusion model demonstrates varying efficacy across different classes. In the Cyst class, its accuracy and reliability in identifying relevant cases yield a score of 0.88. In the Normal class, it achieves a score of 0.93 due to its accurate and comprehensive identification. However, its performance declines in the Stone class, where it reaches a score of 0.66. In the Tumor category, the model shows improvement with a score of 0.88. Overall, it achieves an accuracy rate of 87.77 % which is higher than the two base models. The proposed model indicates the strong results across different classes. For the Cyst class, it achieves a score of 0.93 and an overall accuracy of 91.31 %. In the Normal category, it effectively identifies all relevant cases, achieving an exceptional score of 0.96. In the Stone class, it scores 0.76 due to its high ability to identify relevant cases. In the Tumor class, it excels with scores that reflect precise and comprehensive identification, leading to a strong result of 0.89. Comparing its accuracy to the two base models, the Fusion model, and ConvLSTM with Inception Block, the proposed model outperforms them with an overall accuracy of 91.31 %. This highlights the superior performance of the proposed model in handling multiclass classification tasks on this dataset.

Fig. 10 displays the visualization of the confusion matrix for the proposed model compared to various backbone models, including Fusion model, and ConvLSTM with Inception Block, when applied to a multiclass dataset. Fig. 10(a) and (b) show that both baseline models perform similarly, with VGG-19 slightly outperforming ResNet-50. Fig. 10(c) demonstrates that fusing each model with an Inception block improves performance compared to the individual base models. Furthermore, Fig. 10(d) highlights an additional performance boost when integrating a ConvLSTM unit with the Inception block. Finally, Fig. 10(e) shows that the proposed model outperforms all others, achieving the most accurate classification of 1135 CT images into their respective classes.

4.5. Model interpretability and visualization analysis

In this section, to enhance the assessment of the chosen backbone pre-trained model such as ResNet-50, and VGG19, we conducted a comprehensive study focused on model feature interpretability and visualization throughout the training process [23]. For visualization purposes, we employed GRAD-CAM and Shap interpretability techniques. GRAD-CAM offers valuable insights by highlighting the regions of the image that are crucial for the model predictions, aiding in understanding the model decision-making process and potentially identifying areas for improvement. Fig. 11 (Private Class: Binary), and Fig. 12 (Public Class: Multi) illustrate the GRAD-CAM analysis of backbone model. On the other hand, Shap analysis provides a deeper understanding of feature importance, shedding light on which features contribute most significantly to the model output, thus facilitating model refinement and feature engineering. Fig. 13 (Private Class: Binary), and Fig. 14 (Public Class: Multi) illustrate the Shap analysis of backbone model. Interpretability and visualization techniques are essential for gaining insights into the inner workings of complex models, enabling researchers and practitioners to validate model decisions, detect biases, and ensure model trustworthiness. They also aid in model debugging, facilitating the identification and rectification of erroneous or undesirable behaviors. By employing such techniques, we approve our proposed model the transparency, robustness, and reliability, thereby development trust and confidence in their deployment across various domains.

During model training, it is crucial to analyze the feature extraction process to gain insights into feature learning. For this purpose, many previous research [44,45] efforts have employed well-known techniques like feature map visualization. Feature map analysis offers several benefits. It provides a deeper understanding of how the model transforms input data into meaningful features. Furthermore, it helps in identifying patterns and structures learned by the model, aiding in model interpretability. and, feature map analysis can guide model optimization by revealing areas where the model may be underperforming or overfitting, thus facilitating improvements in overall performance. Fig. 15 (Private Class: Binary), and Fig. 16 (Public Class: Multi) illustrate the feature map analysis of backbone model.

4.6. Model performance evaluation with pre-trained models

We conducted a model performance evaluation using pre-trained models [46,47] to assess their effectiveness in an identical test environment on a new dataset categorized into binary classes such as Normal

and Tumor. This evaluation compares several models across different metrics to identify strengths and weaknesses in their performance. The results in Table 7 provide a comprehensive overview of each model precision, recall, F1-score, and overall accuracy. These insights are crucial for determining the most suitable model for specific applications.

DenseNet201 (Base) model demonstrates varying performance across different evaluation metrics. It achieves a high score of 93.17 for identifying relevant cases effectively, leading to an overall accuracy of 94.05. ResNet101 (Base) model follows closely with a score of 90.36 for relevant case identification and achieves an accuracy of 90.90. ResNet152 (Base) model improves on these metrics slightly, with a score of 91.90 and an accuracy of 92.65. Xception (Base) model also performs well, attaining a strong identification score of 94.30 and an accuracy of 94.08. The proposed model outperforms all with outstanding metrics, a precision of 99.01, a recall of 99.46, an F1-Score of 99.23, and the highest accuracy of 99.30. This demonstrates the superior capabilities of the Proposed Model in this performance evaluation. The outstanding performance of our proposed model can be largely attributed to the implementation of feature fusion techniques. By integrating features from multiple sources or layers, the model gains a more comprehensive representation of the data, which enhances its ability to capture relevant patterns and nuances that might be missed by models using single-source data. This approach not only improves the accuracy and robustness of the model but also significantly enhances its ability to generalize across different and more complex datasets, leading to superior performance metrics as evidenced in the evaluations.

We also conducted a model performance evaluation on a benchmark multiclass dataset that includes classes like Cyst, Normal, Stone, and Tumor. We used different pre-trained models for this assessment. Table 8 details this evaluation by comparing several models. The comparison includes our proposed model and covers metrics like precision, recall, F1-score, and accuracy. Using a variety of pre-trained models allows us to benchmark their performance effectively. This helps us understand which models are most effective for specific data types or classes. Such comparisons are crucial because they highlight each model strengths and weaknesses, which guides further optimizations. By evaluating these models against our proposed model, we demonstrate our approach superior ability to handle complex class distinctions more effectively and accurately. This evaluation not only validates the effectiveness of our proposed model but also deepens our understanding of how different architectures perform in real-world scenarios.

DenseNet201 (Base) model demonstrates moderate effectiveness in various evaluation metrics. It accurately identifies relevant cases with a score of 77.72 and achieves an overall accuracy of 81.49. ResNet101 (Base) model shows slightly lower performance with a score of 77.10 and reaches an accuracy of 83.58. ResNet152 (Base) shows improvement with a score of 79.48 and an accuracy of 85.11. Xception (Base) model performs comparably and delivers a score of 79.41 with an overall accuracy of 82.46. The proposed model significantly surpasses the others by achieving a strong identification score of 92.98 and obtaining the highest overall accuracy of 91.31. This shows the superior capability of the proposed model in handling complex classifications in this dataset.

4.7. Model performance evaluation with existing studies

Table 9 provides a comparative study of model performances across various existing methods and highlighting the accuracy achieved in binary and multi-class classification tasks on same dataset. The author in study [48] achieved 87.90 % accuracy in multi-class classification by using pre-trained ResNet50 model on the Kidney Disease Dataset. In study [49], a weighted average ensemble model achieved an efficient accuracy at 98.75 % in binary classification. The enhanced DL model in study [50] accomplished a 94.09 % accuracy rate in binary classification. The SSLSD-KTD method developed in study [51] scored 98.04 % in binary classification. Our proposed model excelled by achieving the highest accuracy in binary classification at 99.30 % and performed well in multi-class classification with an accuracy of 91.31 %.

5. Conclusion and future work

This study proposed an optimized fusion-based DL model to address key limitations in conventional DL architectures, such as poor generalization, overfitting, and excessive data dependency, specifically for CT scan classification tasks. By integrating ConvLSTM for improved temporal dependency modeling and an inception block for enhanced spatial feature extraction, the model achieved a more comprehensive and robust feature representation. The effectiveness and generalization capability of the proposed model have rigorously evaluated on two diverse datasets. For binary classification, the model achieved an impressive accuracy of 99.6 % on the MHC-CT dataset, effectively distinguishing tumor from normal CT scans. This highlights the model exceptional precision and reliability in detecting abnormalities with minimal error. Additionally, to validate its robustness in more complex scenarios, the model has tested on a publicly available multiclass CT scan dataset, where it attained an accuracy of 91.31 %. This demonstrates the model capability to handle multi-class classification tasks across different imaging modalities, further reinforcing its adaptability in diverse medical imaging applications. The main contributions of this study are as follows:

- **Optimized Fusion-Based Architecture**: The proposed model leverages ConvLSTM and inception block to effectively combine spatial and temporal features, ensuring better generalization compared to conventional DL methods.
- **High-Performance Medical Image Classification**: Achieving 99.6 % accuracy in binary classification and 91.31 % in multiclass classification, the model sets a new benchmark in CT scan analysis, outperforming existing pretrained models.
- **Reduced Overfitting and Enhanced Generalization**: By integrating multiple pretrained architectures, the approach mitigates overfitting issues and reduces dependency on large-scale annotated datasets, making it a more efficient solution for real-world medical applications.
- **Potential for Clinical Applications**: Given its high accuracy and robustness, the model can be a reliable tool for radiologists and healthcare professionals, assisting in automated tumor detection and classification in CT scans.

For future studies, we plan to further validate the proposed model by incorporating additional architectural blocks and optimizing its feature extraction capabilities. Additionally, we aim to evaluate its performance across a broader range of diseases, ensuring its adaptability and robustness in diverse medical imaging applications. Furthermore, we will explore different datasets and real-world clinical scenarios to enhance the model generalizability and practical applicability.

Ethical approval

Approved.

Data availability

The data supporting the findings of this study, along with the source code, private dataset, and models, are available in the GitHub repository. https://github.com/VS-EYE/KidneyDiseaseDetection.git.

Funding

No funding

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] G. Garcia-Garcia, et al., Chronic kidney disease (CKD) in disadvantaged populations, Clinical Kidney J. 8 (1) (2015) 3–6.
[2] C. Elendu, et al., Comprehensive review of current management guidelines of chronic kidney disease, Medicine 102 (23) (2023) e33984.
[3] I. Sorokin, C. Mamoulakis, K. Miyazawa, A. Rodgers, J. Talati, Y. Lotan, Epidemiology of stone disease across the world, World J. Urol. 35 (2017) 1301–1320.
[4] E. Cocchi, J.G. Nestor, A.G. Gharavi, Clinical genetic screening in adult patients with kidney disease, Clin. J. Am. Soc. Nephrol. 15 (10) (2020) 1497–1510.
[5] W.C. O'Neill, Renal relevant radiology: use of ultrasound in kidney disease and nephrology procedures, Clin. J. Am. Soc. Nephrol. 9 (2) (2014) 373–381.
[6] N. Ghaffar Nia, E. Kaplanoglu, A. Nasab, Evaluation of artificial intelligence techniques in disease diagnosis and prediction, Discover Artificial Intelligence 3 (1) (2023) 5.
[7] A. Batool, Y.-C. Byun, Brain tumor detection with integrating traditional and computational intelligence approaches across diverse imaging modalities-Challenges and future directions, Comput. Biol. Med. (2024) 108412.
[8] M. Ayar, A. Isazadeh, F.S. Gharehchopogh, M. Seyedi, Chaotic-based divide-and-conquer feature selection method and its application in cardiac arrhythmia classification, J. Supercomput. (2022) 1–27.
[9] P.K. Pagadala, S.L. Pinapatruni, C.R. Kumar, S. Katakam, L.S.K. Peri, D.A. Reddy, Enhancing lung cancer detection from lung CT scan using image processing and deep neural networks, Rev. Intelligence Artif. 37 (6) (2023).
[10] F. Ahmed, et al., Identification of kidney stones in KUB X-ray images using VGG16 empowered with explainable artificial intelligence, Sci. Rep. 14 (1) (2024) 6173.
[11] V. Yamuna, G. Stalin Babu, G. Vijay Kumar, Y. Manchala, A deep learning framework for kidney stone prediction, in: International Conference on Communications and Cyber Physical Engineering 2018, Springer, 2024, pp. 95–102.
[12] A.S.-M. Hamad, Characterization of Immune Cell Infiltration in Clear Cell Renal Cell Carcinoma, Universitäts-und Landesbibliothek Bonn, 2024.
[13] J.E. Knudsen, J.M. Rich, R. Ma, Artificial intelligence in pathomics and genomics of renal cell carcinoma, Urologic Clinics 51 (1) (2024) 47–62.
[14] R. Nadal, B.P. Valderrama, J. Bellmunt, Progress in systemic therapy for advanced-stage urothelial carcinoma, Nat. Rev. Clin. Oncol. 21 (1) (2024) 8–27.
[15] B. Reuben, C. Narmadha, Effective kidney stone prediction based on optimized Yolov7 segmentation and deep learning classification, Int. J. Intell. Syst. Appl. Eng. 12 (1) (2024) 183–192.
[16] V. Karthikeyan, M.N. Kishore, S. Sajin, End-to-end light-weighted deep-learning model for abnormality classification in kidney CT images, Int. J. Imag. Syst. Technol. 34 (1) (2024) e23022.
[17] N. Sasikaladevi, S. Pradeepa, A. Revathi, S. Vimal, R.G. Crespo, Diagnosis of kidney cyst, tumor and stone from CT scan IMAGESUSING feature FUSION HYPERGRAPH convolutional neural network (F 2 HCN 2), Int. J. Multiscale Comput. Eng. 22 (5) (2024).
[18] M.S. Farooq, A. Tariq, A Deep Learning Architectures for Kidney Disease Classification, 2024 *arXiv preprint arXiv:2403.15895*.
[19] T. Hossain, F. Sayed, S. Islam, Adaptive Local Binary Pattern: A Novel Feature Descriptor for Enhanced Analysis of Kidney Abnormalities in CT Scan Images Using Ensemble Based Machine Learning Approach, 2024 *arXiv preprint arXiv: 2404.14560*.
[20] D.M. Alsekait, et al., Toward comprehensive chronic kidney disease prediction based on ensemble deep learning models, Appl. Sci. 13 (6) (2023) 3937.
[21] L.B. Sorkhabi, F.S. Gharehchopogh, J. Shahamfar, A systematic approach for pre-processing electronic health records for mining: case study of heart disease, Int. J. Data Min. Bioinf. 24 (2) (2020) 97–120.
[22] M. Ayar, A. Isazadeh, F.S. Gharehchopogh, M. Seyedi, NSICA: multi-objective imperialist competitive algorithm for feature selection in arrhythmia diagnosis, Comput. Biol. Med. 161 (2023) 107025.
[23] S.U.R. Khan, S. Asif, O. Bilal, S. Ali, Deep hybrid model for Mpox disease diagnosis from skin lesion images, Int. J. Imag. Syst. Technol. 34 (2) (2024) e23044.
[24] Z. Khan, M.Z. Hossain, N. Mayumu, F. Yasmin, Y. Aziz, Boosting the prediction of brain tumor using two stage BiGait architecture, in: 2024 International Conference on Digital Image Computing: Techniques and Applications (DICTA), IEEE, 2024, pp. 411–418.
[25] U.S. Khan, S.U.R. Khan, Boost diagnostic performance in retinal disease classification utilizing deep ensemble classifiers based on OCT, Multimed. Tool. Appl. (2024) 1–21.
[26] A. Raza, M.T. Meeran, U. Bilhaj, Enhancing breast cancer detection through thermal imaging and customized 2D CNN classifiers, VFAST Trans. Software Eng. 11 (4) (2023) 80–92.
[27] F.A. Özbay, E. Özbay, F.S. Gharehchopogh, An improved artificial rabbits optimization algorithm with chaotic local search and opposition-based learning for engineering problems and its applications in breast cancer problem, CMES-Comput. Model. Eng.; Sci. 141 (2) (2024).
[28] K. Yildirim, P.G. Bozdag, M. Talo, O. Yildirim, M. Karabatak, U.R. Acharya, Deep learning model for automated kidney stone detection using coronal CT images, Comput. Biol. Med. 135 (2021) 104569.
[29] I. Aksakalli, S. Kaçdioğlu, Y.S. Hanay, Kidney x-ray images classification using machine learning and deep learning methods, Balkan J. Electr. Comput. Eng. 9 (2) (2021) 144–151.
[30] S.D. Mahalakshmi, An optimized transfer learning model based kidney stone classification, Comput. Syst. Sci. Eng. 44 (2) (2023).
[31] S. Sudharson, P. Kokil, Computer-aided diagnosis system for the classification of multi-class kidney abnormalities in the noisy ultrasound images, Comput. Methods Progr. Biomed. 205 (2021) 106071.
[32] R.A. Dos Santos, Employing advanced deep learning technology for the detection of kidney stones in unenhanced computed tomography (CT) imaging: a model-based approach, Int. J. Technol., Innov. Manag.(IJTIM) 3 (2) (2023) 16–21.
[33] Jyotismita Chaki, Ayşegül Uçar, An efficient and robust approach using inductive transfer-based ensemble deep neural networks for kidney stone detection, IEEE Access 12 (2024) 32894–32910.
[34] Ö. Sabuncu, B. Bilgehan, E. Kneebone, O. Mirzaei, Effective deep learning classification for kidney stone using axial computed tomography (CT) images, Biomed. Eng./Biomed. Technik 68 (5) (2023) 481–491.
[35] P. Yadav, S. Sharma, HFBO-KSELM: hybrid flash butterfly optimization-based kernel softplus extreme learning machine for classification of chronic kidney disease, J. Supercomput. 79 (15) (2023) 17146–17169.
[36] Y. Wu, Z. Yi, Automated detection of kidney abnormalities using multi-feature fusion convolutional neural networks, Knowl. Base Syst. 200 (2020) 105873.
[37] P.K. Rao, S. Chatterjee, K. Nagaraju, S.B. Khan, A. Almusharraf, A.I. Alharbi, Fusion of graph and tabular deep learning models for predicting chronic kidney disease, Diagnostics 13 (12) (2023) 1981.
[38] I. Shahzad, S.U.R. Khan, A. Waseem, Z.U. Abideen, J. Liu, Enhancing ASD classification through hybrid attention-based learning of facial features, Signal, Image and Video Processing (2024) 1–14.
[39] M.N. Islam, M. Hasan, M.K. Hossain, M.G.R. Alam, M.Z. Uddin, A. Soylu, Vision transformer and explainable transfer learning models for auto detection of kidney cyst, stone and tumor from CT-radiography, Sci. Rep. 12 (1) (2022) 1–14.
[40] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[41] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, 2014 *arXiv preprint arXiv:1409.1556*.
[42] K.K. Mohbey, S. Sharma, S. Kumar, M. Sharma, COVID-19 identification and analysis using CT scan images: deep transfer learning-based approach, in: Blockchain Applications for Healthcare Informatics, Elsevier, 2022, pp. 447–470.
[43] Y. Tatsunami, M. Taki, Sequencer: deep lstm for image classification, Adv. Neural Inf. Process. Syst. 35 (2022) 38204–38217.
[44] S.U.R. Khan, M. Zhao, S. Asif, X. Chen, Hybrid-NET: a fusion of DenseNet169 and advanced machine learning classifiers for enhanced brain tumor diagnosis, Int. J. Imag. Syst. Technol. 34 (1) (2024) e22975.
[45] S.U.R. Khan, M. Zhao, Y. Li, Detection of MRI brain tumor using residual skip block based modified MobileNet model, Clust. Comput. 28 (4) (2025/02/25 2025) 248, https://doi.org/10.1007/s10586-024-04940-3.
[46] S.U.R. Khan, S. Asif, M. Zhao, W. Zou, Y. Li, Optimize brain tumor multiclass classification with manta ray foraging and improved residual block techniques, Multimed. Syst. 31 (1) (2025) 1–27.
[47] S.U.R. Khan, S. Asif, M. Zhao, W. Zou, Y. Li, X. Li, Optimized deep learning model for comprehensive medical image analysis across multiple modalities, Neurocomputing (2024) 129182.
[48] M.S. Hossain, S. Hassan, M. Al-Amin, R. Hossain, Kidney Disease Detection and Classification from CT Images Using Watershed Segmentation and Deep Learning, Brac University, 2022.
[49] S. Srivastava, R.K. Yadav, V. Narayan, P.K. Mall, An ensemble learning approach for chronic kidney disease classification, J. Pharm. Negat. Results (2022) 2401–2409.
[50] M. Majid, et al., Enhanced transfer learning strategies for effective kidney tumor classification with CT imaging, Int. J. Adv. Comput. Sci. Appl. 14 (2023) 2023.
[51] E. Özbay, F.A. Özbay, F.S. Gharehchopogh, Kidney tumor classification on CT images using self-supervised learning, Comput. Biol. Med. (2024) 108554.