Combining convolutional operators in unsupervised networks for kidney abnormalities

Aekkarat Suksukont[1], Anuruk Prommakhot[2], Jakkree Srinonchat[3]
[1]Department of Computer Engineering, Faculty of Industrial Education, Rajamangala University of Technology Suvarnabhumi, Suphanburi, Thailand
[2]Department of Computer Engineering, Faculty of Engineering, Rajamangala University of Technology Thanyaburi, Pathum Thani, Thailand
[3]Department of Electronics and Telecommunication Engineering, Faculty of Engineering, Rajamangala University of Technology Thanyaburi, Pathum Thani, Thailand

Article Info	**ABSTRACT**

Deep learning plays a pivotal role in advancing the diagnosis of renal dysfunction, achieving performance levels comparable to those of medical experts. However, disease domain variations and model differences can impact learning quality. To address renal dysfunction, we propose dual-stream convolutional (DSC) and dual-input convolutional (DIC) for unsupervised learning. The proposed network is designed to process multi-scale data and employs parallel data aggregation to enhance learning capabilities, improving the reliability of the experimental results. DSC achieved training losses of 0.0069, 0.0056, 0.0042, and 0.0048 for normal, cyst, stone, and tumor datasets, respectively, while DIC achieved losses of 0.0066, 0.0063, 0.0044, and 0.0058 for the same categories. The experimental results demonstrate that our proposed models outperform state-of-the-art approaches, making them well-suited for broad application in clinical research studies.

Corresponding Author:

Jakkree Srinonchat
Department of Electronics and Telecommunication Engineering, Faculty of Engineering
Rajamangala University of Technology Thanyaburi
Pathum Thani, Thailand
Email: jakkree.s@en.rmutt.ac.th

1. INTRODUCTION

Deep learning has gained wide acceptance across various research fields, particularly in the medical domain, due to its ability to collect data and learn complex features or patterns for detection, segmentation, classification, and recognition, especially in diagnosing diseases through magnetic resonance imaging (MRI), X-ray, or microscopic images [1]–[3]. While deep learning yields highly efficient results, understanding the essential components of disease, such as regions of interest or disease positions in medical images, remains crucial for explaining disease hallmarks and the network's ability to learn the pathogenesis [4]. Overlapping features, which obscure critical disease components [5], and noise in the images [6], [7] are significant factors that impact learning and can lead to noise signals before errors occur. These challenges, mainly when dealing with networks tasked with learning the identity or patterns of overlapping diseases, such as COVID-19 and kidney abnormalities in X-ray images [8], [9], or brain tumors and kidney disease in MRI scans [10], [11], pose limitations that require careful attention.

In recent years, deep learning techniques have been developed to enhance the efficiency of learning complex disease components and handling disease features. These advancements have led to notable

performance improvements, such as enhancing convolutional neural network (CNN) parallelization with residual fusion to mitigate vanishing gradient issues and applying DropBlock techniques to better extract key features from complex medical images [12]. The integration of visual geometry group (VGG)-16 and VGG-19 has enhanced the ability to capture both shallow and deep features, thereby improving medical insight detection [13].

The development of the adaptive hybridized deep CNN focuses on residual learning and dilated convolutions to address missing information, allowing for broader and more detailed feature extraction. It incorporates autoencoders and skip connections to pass information to deeper layers and reduce the dimensionality of complex image data for more accurate disease recognition in large images [14]. A modified deep belief network has also been designed as a classification algorithm for predicting kidney-related diseases, using Softmax as the activation function and categorical cross-entropy as the loss function for early detection of chronic diseases [15].

The you only look once version 8 (YOLOv8) model has been employed to analyze differentiated medical features, demonstrating its effectiveness in disease detection and diagnosis [16]. Additionally, deep learning has been applied to address the challenge of disease characteristics that arise from the diversity of image data, such as using multi-scale learning to handle varying data complexities. Networks are often designed with different receptive fields or convolutional operators (multi-scale) to capture target patches effectively [17], [18], providing significant design advantages over traditional machine learning methods.

In previous works [12]–[18], they investigated the effectiveness of deep learning approaches in addressing challenges related to imaging and improving learning outcomes from imbalanced medical data. Although deep learning has demonstrated outstanding potential for medical image applications, some critical issues remain unresolved. First, image quality impacts overall learning performance, or changes in the data domain influence the quality of the network's learning. Second, designing CNNs to handle different data dimensions affects the network's ability to capture features.

These issues are central to the development of deep learning, highlighting the need for continued improvement in deep learning techniques. We propose a learning model for kidney disorders using a convolutional operator fusion strategy with VGGNet, which is divided into two models: dual-stream convolutional (DSC), designed to capture additional features and address information loss during feature extraction and dual-input convolutional (DIC), which is designed to learn multi-scale data and capture complex features, allowing the model to adapt effectively to constrained domains. The proposed network leverages unsupervised learning to manage complex data and reduce the dimensionality of medical images.

2. METHOD

2.1. Convolutional operator

The convolutional operator [19] is an essential mathematical function that extracts features from input data. It highlights crucial patterns, such as edges, corners, or textures. The operator works with a kernel, which slides across the input data, performing convolutions at each position to detect features like edges and textures. Stride, the step size of the filter's movement, reduces the output size to minimize computational complexity.

The same padding enhances this process, ensuring the output size matches the input size. The rectified linear unit (ReLU) activation function is also applied after convolution to remove negative values, allowing the model to learn more complex patterns. As mentioned, the convolutional operator in CNNs works by multiplying pixel values from the input with the filter's weights, combining the results into a single value at each position of the feature map, as in (1).

$$S_{(i,j)} = (I * K)_{(i,j)} \sum_m \sum_n I_{(i+m,j+m)} \cdot K_{(m,n)} \qquad (1)$$

Where $S_{(i,j)}$ represents the output at position i, j on the feature map, or the result of the convolution, (i, j) is the input value at position i, j on the input image, and $K_{(m,n)}$ denotes the weights in the filter or kernel at position m, n. The symbol * represents the convolution operation.

2.2. Dual convolutional operator

Dual convolutional [20] operators are convolutional filters designed to extract features from different data viewpoints. Each filter is tailored to detect specific features, such as horizontal edges, vertical edges, or complex surfaces, and these features are then combined into a single feature map. This approach enables the model to capture in-depth information and features of varying sizes. In this study, both filters are concatenated into a vector to form the final feature map, as in (2).

$$S_{combined(i,j)} = S_{1(i,j)} + S_{2(i,j)} \qquad (2)$$

Where $S_{1(i,j)}$ is the output of the first convolution at position (i,j), $S_{2(i,j)}$ is the output of the second convolution operation at the same position (i,j), and $S_{combined(i,j)}$ is the sum of these two values at each position (i,j). This study concatenates feature maps into a vector to form the final feature map.

2.3. Residual layers

Residual layers [21] are designed to address the problem of vanishing gradients, which often occurs as networks deepen, by introducing skip connections. These connections allow the input of a given block to bypass the transformations in other layers and be directly added to the output, enabling more effective learning of new features, as shown in (3).

$$y = f(x) + x \qquad\qquad (3)$$

Where x is the input, $f(x)$ represents the processing function of the convolutional layer, and the final result is obtained by adding the values of x and $f(x)$, residual layers also enables the model to learn more complex transformations while maintaining its trainability.

2.4. Deep learning techniques

Deep learning involves networks that process data through multiple layers, with each layer learning progressively complex features. This research employs VGG-19 [22], a deep learning architecture designed to capture intricate features. The network is notable for its use of small filters and a consistent structure across all layers, a key feature that enhances its ability to collect features from diverse data and improve overall performance efficiently.

3. PROPOSED METHOD

This research aims to classify renal abnormalities in computed tomography (CT) scan images. Our experiments utilize the well-established VGGNet, incorporating loosely combined convolutional operators to enhance its ability to capture variance-dependent features. Unlike the original deep-structured expanded networks, our approach introduces a distinct experimental framework. The techniques and experiments are divided into three parts, as shown in Figure 1. In Figure 1(a), this network combines a modified convolutional operator (MCO) with VGG-19 to improve learning from medical image data. Input images with a size of 100×100×3 pixels are fed into both branches. On the MCO side, the network uses convolutional kernels with 32, 64, 128, and 256 filters, each with a 3×3 kernel, SAME padding, and a ReLU activation function.

These layers are grouped into blocks consisting of three stacked layers, followed by a 2×2 max-pooling layer at the end of each block. Residual connections are added to help avoid the vanishing gradient problem as the network deepens. In parallel, the VGG-19 branch applies its standard convolutional architecture, using multiple convolutions and max-pooling layers to extract deep feature representations. After processing, the features from both branches are fused, flattened into a vector, and passed through three fully connected layers with 120, 500, and 500 nodes, respectively, before reaching the final output layer. This model design focuses on improving feature fusion and adaptability to complex data by learning diverse features. The network was trained on the kidney dataset, as in Figure 2, using the hyperparameters as in Table 1.

Table 1. Parameters for training the model

Parameter	Value
Input size	100×100×3, 150×50×3, 200×200×3
Max epochs	20
β1 and β2 (Adam optimized)	0.9, 0.0009
Function	Sigmoid
Epsilon	1e-8
Batch size	2
Learning rate	10^{-3}, 10^{-4}, 10^{-5}

In Figure 1(b), the design takes dual-input data with different characteristics. Images sized 150×150×3 pixels are fed into the MCO, which uses a convolutional kernel with 32, 64, 128, and 256 filters. Each layer applies a 3×3 kernel with SAME padding and ReLU activation, arranged into blocks of three stacked layers followed by a 2×2 max-pooling layer. Residual connections also support stable learning and reduce the risk of vanishing gradients. Meanwhile, the second branch receives 200×200×3 pixel images and processes them through the standard VGG-19 architecture, which extracts deep features using 3×3 kernels, 2×2 max-pooling layers, and convolutional kernels from 64 to 512. Once both streams complete feature extraction, the outputs are fused and passed through two fully connected layers, each with 500 nodes, before reaching the final output layer. This architecture is designed to process multiple image features

simultaneously, making it highly efficient in handling complex datasets. The network was trained on the kidney dataset as in Figure 2, using the parameters in Table 1.
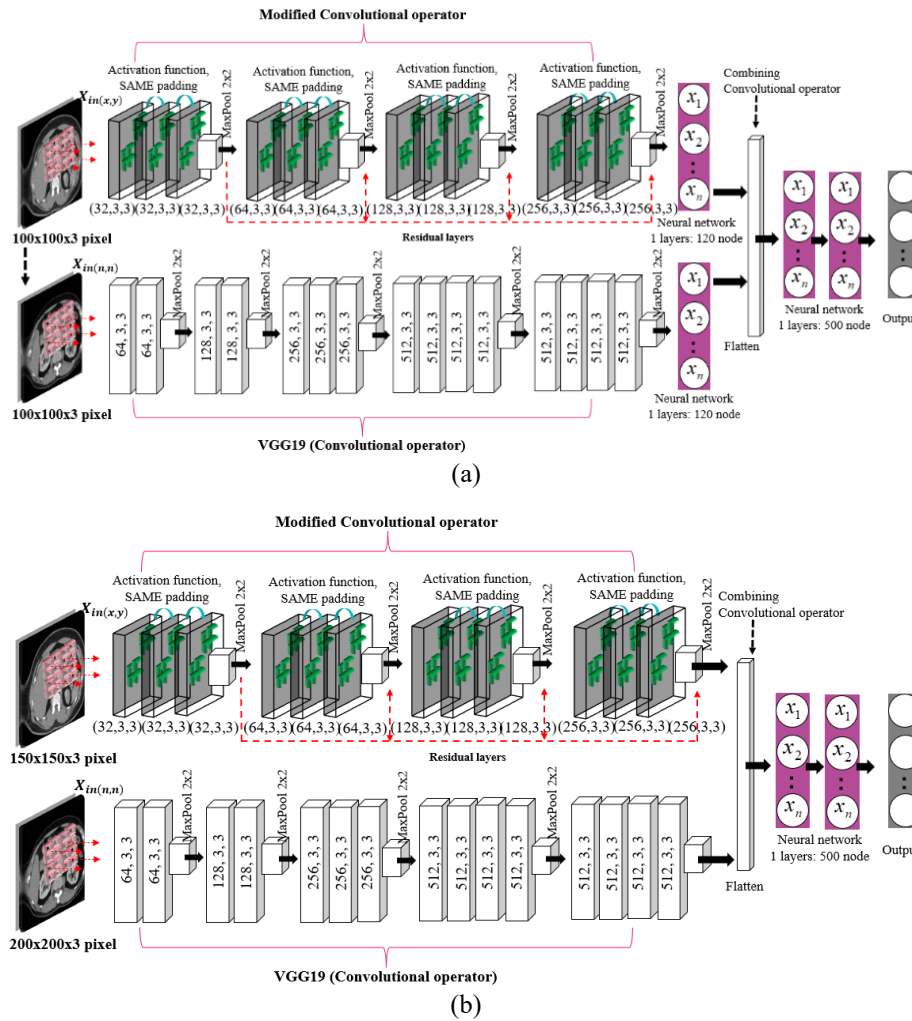

(a)

(b)

Figure 1. Proposed method for kidney abnormality of (a) DSC fusion for unsupervised learning and (b) DIC fusion for unsupervised learning

4. EXPERIMENTAL AND RESULTS

4.1. Kidney dataset

A kidney imaging dataset was compiled from the picture archiving and communication system (PACS) of various hospitals in Dhaka, Bangladesh [13], [23]. The dataset consists of images from patients diagnosed with tumors, cysts, stones, or normal conditions. After thorough examination, 12,446 unique images were generated, including 5,077 normal images as shown in Figure 2(a), 3,709 cyst images as shown in Figure 2(b), 1,377 stone images as shown in Figure 2(c), and 2,283 tumor images as shown in Figure 2(d). The dataset was uploaded to Kaggle to encourage other researchers to develop innovative methods. It was split into 80% for the training set, 10% for the validation set, and 10% for the test set for benchmarking purposes.
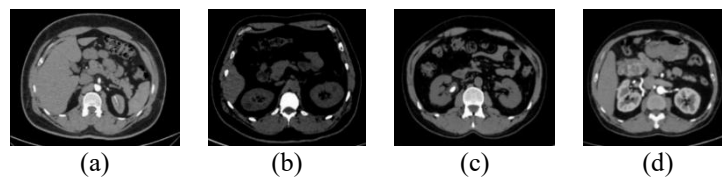

(a)   (b)   (c)   (d)

Figure 2. Kidney abnormalities dataset of (a) normal, (b) cyst, (c) stone, and (d) tumor

4.2. Experiment setup

The setup consists of a Windows system with an Intel Core i5-12400F LGA 1700 CPU @ 2.5 GHz, 32 GB of RAM (clocked at 5,600 MHz), and an NVIDIA RTX 4070 GPU with 12 GB of VRAM and 5,888 CUDA cores. The code is implemented using NumPy and TensorFlow. The proposed model and deep learning network are also trained using standardized parameters, as summarized in Table 1.

4.3. Evaluation

The mean squared error (MSE) evaluates the difference between predicted and actual values, measuring their deviation and always yielding a positive result. In this study, it is used to assess the capability of the unsupervised network, as shown in (4), where n is the number of samples. y_i is the actual value of the sample i, $\hat{y}_i$ is the predicted value of the sample i, $(y_i - \hat{y}_i)^2$ is the squared error of the sample i.

$$MSE = \frac{1}{2n} + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{4}$$

4.4. Performance of training model

This study presents a system for detecting kidney abnormalities, focusing on networks with dense layers and dual input scales based on unsupervised learning. The experiments are conducted on a dataset containing four types of abnormalities: normal, cyst, stone, and tumor. MSE is employed to evaluate the learning performance of the proposed experiments.

Figure 3 shows experimental results using dual dense layers with different hyperparameter settings. Observing the learning process at each epoch with learning rates of 10^{-3}, 10^{-4}, 10^{-5}, we notice rapid learning that points to stable performance. In Figure 3(a), normal has the lowest MSE at 0.0069, which is due to the better performance of the 10^{-4} compared to 10^{-3} and 10^{-5}, which are 0.0093 and 0.0106, respectively. In Figure 3(b), the cyst achieves the lowest MSE of 0.0056 with 10^{-4}, which is higher than 10^{-3} and 10^{-5} by 0.0028 and 0.006, indicating that 10^{-4} provides better learning stability. Similarly, in Figure 3(c), stone records the lowest MSE at 0.0042 using the 10^{-4}, reflecting its effective capability in reducing loss. Finally, in Figure 3(d), the tumor achieves the lowest MSE at 0.0048 with the 10^{-4}. These results show the overall effectiveness and consistency of the 10^{-4} across all data types. Furthermore, across all conditions (normal, cyst, stone, and tumor), the learning rate of 10^{-4} offers the most balanced results, achieving the lowest final loss in MSE with a training speed of 20 epochs. The higher learning rate of 10^{-3} shows faster convergence but slightly higher loss, while the lower learning rate of 10^{-5} results in slower learning and the highest loss in all experiments.
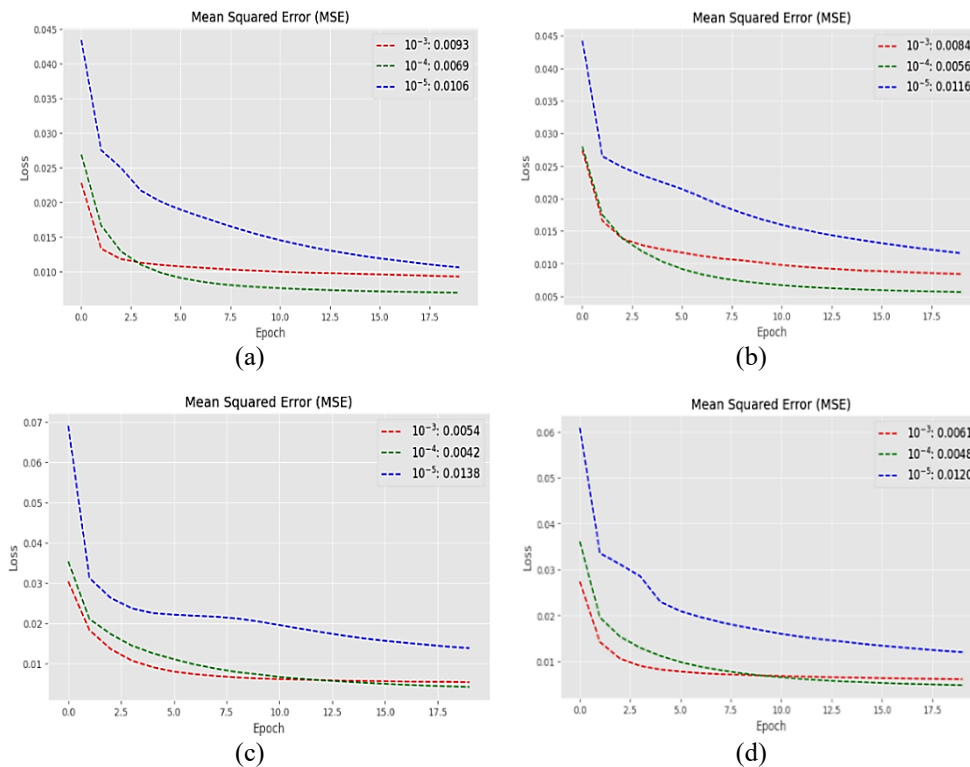

Figure 3. Loss of DSC fusion for unsupervised learning of (a) normal, (b) cyst, (c) stone, and (d) tumor

Figure 4 shows experimental results using dual input scales with different hyperparameter settings. Observing the learning at each iteration for learning rates of 10^{-3}, 10^{-4}, and 10^{-5}, we see fast learning across all rates, indicating effective learning. In Figure 4(a), normal has the lowest MSE at 0.0066, which is due to the better 10^{-4} compared to 10^{-3} and 10^{-5}, which are 0.0090 and 0.0136, respectively. In Figure 4(b), cyst has the lowest MSE from 10^{-4} with a value of 0.0063, which is higher than 10^{-3} and 10^{-5} by 0.0016 and 0.007, indicating that 10^{-4} has better learning stability. Similarly, in Figure 4(c), stone has the lowest MSE at 0.0044 from the experiment with 10^{-4}, demonstrating its effective loss reduction capability. Finally, in Figure 4(d), tumor has the lowest MSE at 0.0058, which is due to the 10^{-4}. The experimental results confirm the overall effectiveness of this learning rate across all data types. Additionally, the curve for 10^{-4} has a higher learning rate than that of 10^{-3} which shows faster learning but with a slightly higher loss across all experiments. At 10^{-5}, the curve progresses more slowly but achieves the lowest loss values in every experiment.
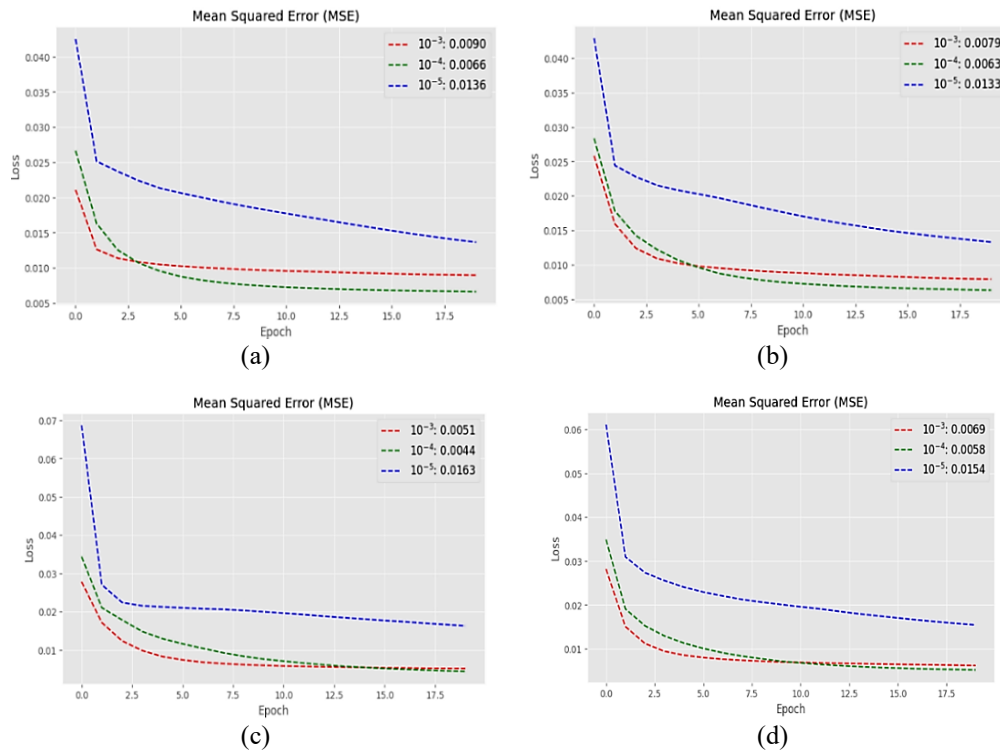

Figure 4. Loss of DIC fusion for unsupervised learning of (a) normal, (b) cyst, (c)stone, and (d) tumor

4.5. Density analysis

Principal component analysis (PCA) on a heatmap is a technique used in TensorFlow to describe the density of renal abnormalities. This study analyzes the differences between four conditions: normal, cysts, stones, and tumors. The component description utilizes the concatenated layer results to demonstrate the proposed network's capability for unsupervised feature fusion. The data density, representing changes and distribution in each condition, is visualized in Figures 5 and 6.

In Figure 5, the composition of the normal, cyst, stone, and tumor data reveals significant differences in density distribution. The normal data, as in Figure 5(a), shows two distinct density peaks, clustered between PC1: -45 and 45, with moderate variability along PC2 and a wide dispersion. In contrast, the cyst data, as shown in Figure 5(b), present a highly concentrated cluster between PC1: -45 and 60, with a relatively low-density peak, indicating uniformity in the cyst profile. The stone data, as in Figure 5(c), shows two distinct peaks, one centered around PC1: -60 and 80, suggesting distinct subgroups within the stone data. Lastly, as in Figure 5(d), the tumor data exhibit the highest variability, with a broad dispersion along both PC1 and PC2, reflecting the complexity of tumor characteristics. These differences in density and distribution underscore the varying levels of variation, with cyst being the most uniform and tumor the most diverse.

Figure 6 shows the data density and distribution differences across the principal components. The normal data, as in Figure 6(a), exhibits two peaks spread from PC1: -15 to 30, reflecting a balanced distribution. As in Figure 6(b), the cyst data show the least spread, with contours concentrated around PC1: -25 to 40 and a lower density peak, indicating homogeneity. The stone data, as in Figure 6(c), spans a

broader range from PC1: -60 to 80 and shows two distinct peaks, suggesting a moderate level of variability. In contrast, as in Figure 6(d), the tumor data displays the widest spread, ranging from PC1: -45 to 60, with three high-density peaks indicating high variability. These results indicate that numerical and density distributions are key indicators of the complexity and variability of each condition.
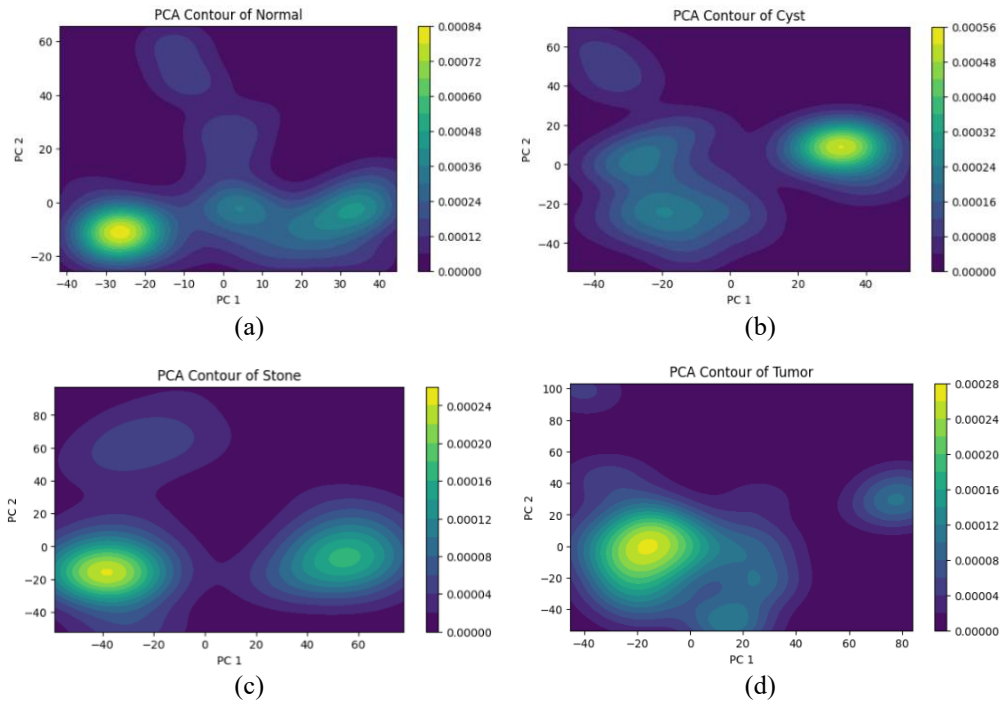

Figure 5. Feature map visualization of DSC fusion for unsupervised learning of (a) density of concatenate layers (normal (10^{-4})), (b) density of concatenate layers (cyst (10^{-4})), (c) density of concatenate layers (stone (10^{-4})), and (d) density of concatenate layers (tumor (10^{-4}))
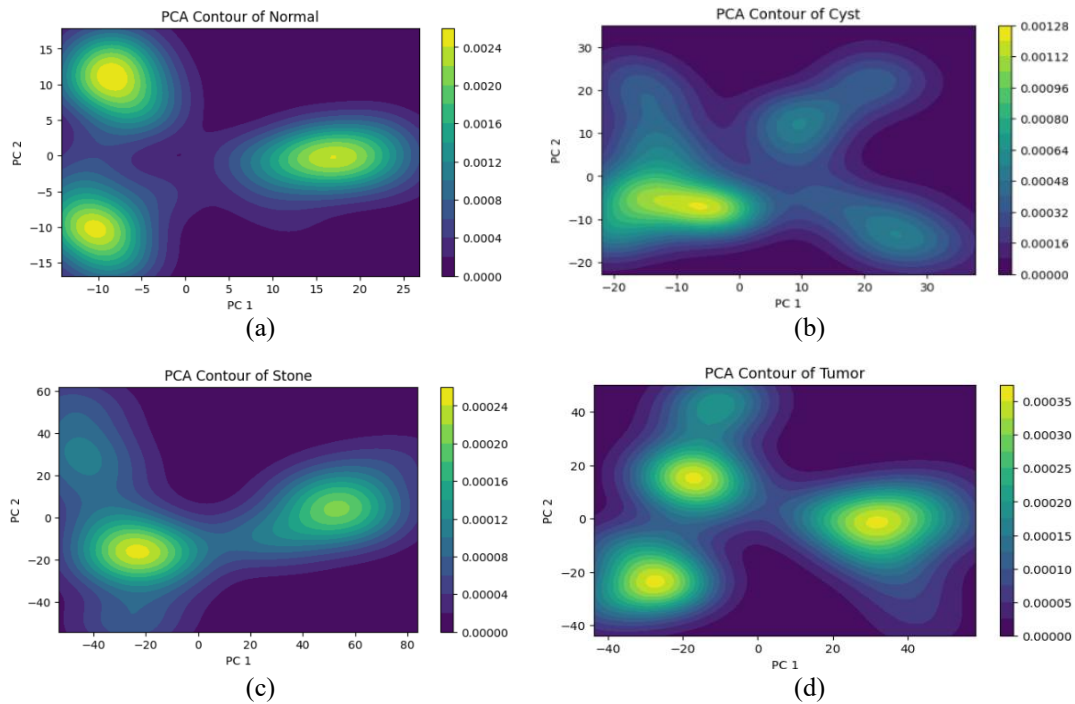
Figure 6. The density of concatenate layers from DIC fusion for unsupervised learning of (a) density of concatenate layers (normal (10^{-4})), (b) density of concatenate layers (cyst (10^{-4})) (c) density of concatenate layers (stone (10^{-4})), and (d) density of concatenate layers (tumor (10^{-4}))

4.6. Comparison model

In this section, we select other state-of-the-art models (VGG-16, VGG-19, Xception, DenseNet-121, and EfficientNetB7 [22], [24]–[26]), which are based on unsupervised learning and adapted from previous studies on kidney abnormalities, to compare their performance with the DSC and DIC models. The parameters used for the experiments are shown in Table 1. The experimental results are summarized in Figure 7.

Figure 7 compares the loss values across different models: VGG-16, VGG-19, Xception, DenseNet-121, and EfficientNetB7 on the normal, cyst, stone, and tumor datasets. The results indicate that EfficientNetB7 demonstrates the best training performance across all groups, achieving losses of $4.3433e^{-04}$, 0.0010, 0.0033, and 0.0015 for the normal, cyst, stone, and tumor datasets, as in Figures 7(a) to 7(d). DenseNet-121 and Xception show slightly higher losses than EfficientNetB7, as in Figures 7(a) to 7(d), but they still efficiently learn the data. In contrast, VGG-16 and VGG-19, as in Figures 7(a) to 7(d), exhibit higher loss values than the other models, with significantly slower learning rates. This suggests that these networks struggle with more complex tasks and require further tuning. Compared to the proposed methods in Figures 3 and 4, it is evident that while DSC and DIC have lower performance than EfficientNetB7, they are comparable to DenseNet-121 and Xception, and outperform VGG-16 and VGG-19.
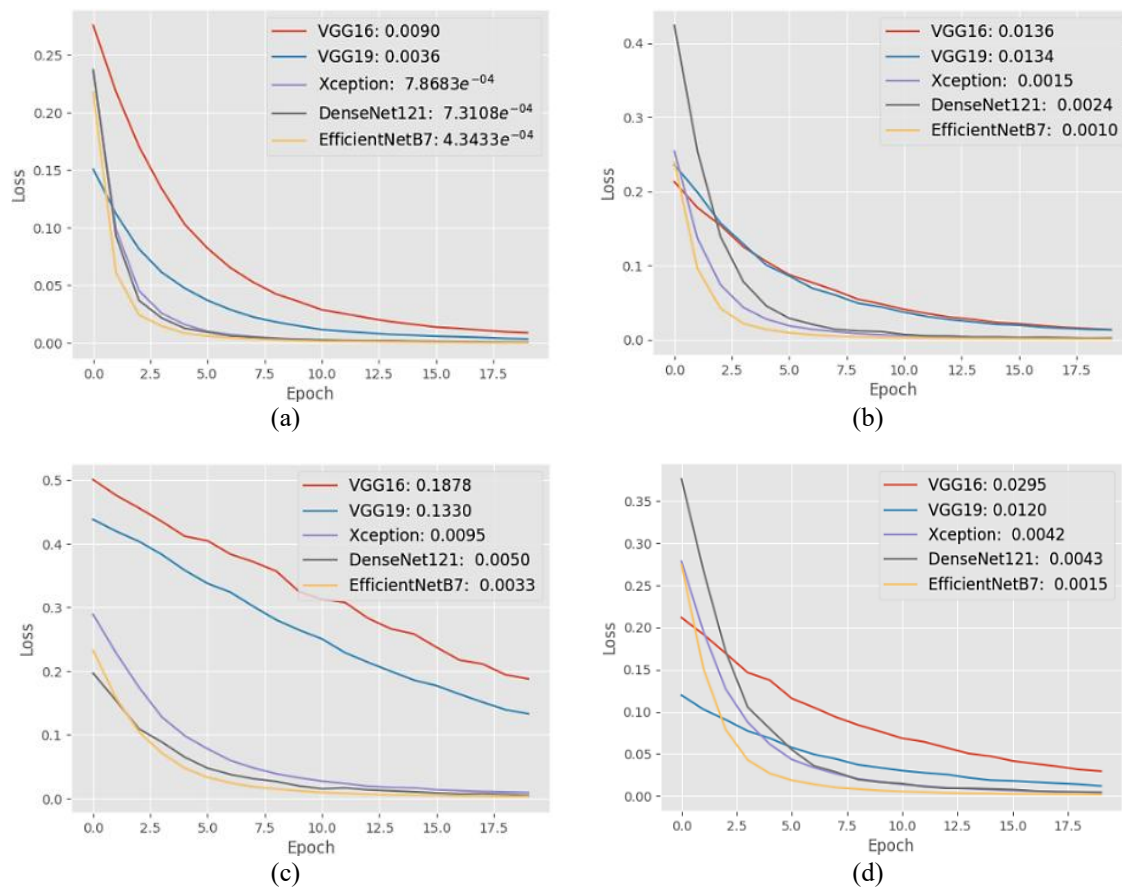

Figure 7. Loss comparison across various models for kidney abnormality of (a) normal (10^{-5}), (b) cyst (10^{-5}), (c) stone (10^{-5}), and (d) tumor (10^{-5})

5. DISCUSSION

This study proposes a combination of a CNN and the VGG-19 in a DSC network to enhance its performance and capability in learning kidney-medical features. The model design in Figure 1(a) employs identical input dimensions (100×100×3) for a two-stream network [27]. It integrates different 32, 64, 128, and 256 convolutional kernels and 2×2 kernel sizes to capture multi-scale spatial features and their dependencies effectively. Additionally, residual connections between processing blocks help mitigate the vanishing gradient problem and facilitate the learning of deep networks. In contrast, Figure 1(b) shows the two different input sizes (150×150×3 and 200×200×3) to improve the network's ability to process images with varying characteristics. This DIC approach [28], [29] enables the model to learn from multiple data

sources, thereby enhancing its deep feature learning capacity and facilitating a comparative analysis of responses across input variations.

Our experiments with the DSC model suggest it achieves a reasonable balance between complexity and performance, as measured by MSE. In the experiments, a learning rate of 10^{-4} yields the lowest loss across all classes shown in Figures 3(a) to 3(d), namely normal, cyst, stone, and tumor. The obtained accuracies reflect the model's capability to learn complex features from medical data effectively. Although a higher learning rate of 10^{-3} leads to faster convergence, it results in a higher loss. Conversely, a lower learning rate of 10^{-5} slows down learning and fails to reduce the loss as effectively of 10^{-4} within the given number of epochs. These results indicate that the proposed network design enhances the model's learning capability and is suitable for analyzing various types of medical image data.

The experimental results and studies in this research show that the proposed DSC and DIC models, which combine the functionality of convolutional operators and VGG-19 networks, are effective in learning complex medical image features, especially when using an appropriate learning rate of 10^{-4}, which yields the lowest loss across all data groups. Although the proposed models are inferior to EfficientNetB7 [26] in terms of achieving the lowest loss, they demonstrate learning performance comparable to advanced models such as Xception [24] and DenseNet-121 [25], and outperform VGG-16 and VGG-19 [22]. In addition, the dual-input stream design improves the ability to process data with different characteristics. However, this research also identifies some limitations that should be addressed in future work, such as reducing network complexity, applying regularization techniques to prevent overfitting, incorporating a variety of clinical indicators, exploring transfer learning [30] approaches, and designing lightweight models [31] for effective application in real-world medical devices.

6. CONCLUSION

This paper presents a method for combining convolutional operators in unsupervised networks for kidney abnormalities. The proposed network integrates two types of operators: i) DSC, which extracts features from a single input through parallel pathways and ii) DIC, which processes images of varying resolutions to enhance multi-scale learning. This enables the network to learn multi-view features and helps reduce overfitting during training. The technique is tested on the kidney abnormalities dataset, a widely accepted and publicly available resource for research and development. Experimental results show that the proposed network outperforms others with a lower training loss rate. This success demonstrates that the method is effective and opens up opportunities for future work, particularly in applying multi-scale and multi-input models to enhance 3D disease detection in medical imaging.

FUNDING INFORMATION

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Aekkarat Suksukont	✓	✓	✓	✓		✓		✓	✓	✓				
Anuruk Prommakhot	✓	✓	✓		✓	✓				✓				
Jakkree Srinonchat	✓		✓		✓	✓	✓		✓	✓	✓	✓	✓	✓

C : **C**onceptualization	I : **I**nvestigation	Vi : **Vi**sualization
M : **M**ethodology	R : **R**esources	Su : **Su**pervision
So : **So**ftware	D : **D**ata Curation	P : **P**roject administration
Va : **Va**lidation	O : Writing - **O**riginal Draft	Fu : **Fu**nding acquisition
Fo : **Fo**rmal analysis	E : Writing - Review & **E**diting	

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are openly available in the CT KIDNEY dataset: normal-cyst-tumor and stone at https://www.kaggle.com/datasets/nazmul0087/ct-kidney-dataset-normal-cyst-tumor-and-stone/data.

REFERENCES

[1] Z. Xu, Y. Liu, G. Xu, and T. Lukasiewicz, "Self-supervised medical image segmentation using deep reinforced adaptive masking," *IEEE Transactions on Medical Imaging*, vol. 44, no. 1, pp. 180–193, 2025, doi: 10.1109/TMI.2024.3436608.

[2] L. Fan, M. Li, Z. Hu, Y. Hong, and D. Kong, "DNGG: Medical image lossless encryption via deep network guided generative," *IEEE Signal Processing Letters*, vol. 32, pp. 1331–1335, 2025, doi: 10.1109/LSP.2025.3552528.

[3] T. Dhar, N. Dey, S. Borra, and R. S. Sherratt, "Challenges of deep learning in medical image analysis—improving explainability and trust," *IEEE Transactions on Technology and Society*, vol. 4, no. 1, pp. 68–75, 2023, doi: 10.1109/tts.2023.3234203.

[4] J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang, "Segment anything in medical images," *Nature Communications*, vol. 15, no. 1, 2024, doi: 10.1038/s41467-024-44824-z.

[5] X. Li *et al.*, "Deep learning based de-overlapping correction of projections from a flat-panel micro array X-ray source: simulation study," *Physica Medica*, vol. 111, 2023, doi: 10.1016/j.ejmp.2023.102607.

[6] S. H. Abbood, H. N. A. Hamed, M. S. M. Rahim, A. Rehman, T. Saba, and S. A. Bahaj, "Hybrid retinal image enhancement algorithm for diabetic retinopathy diagnostic using deep learning model," *IEEE Access*, vol. 10, pp. 73079–73086, 2022, doi: 10.1109/ACCESS.2022.3189374.

[7] Y. Luo *et al.*, "Dehaze of cataractous retinal images using an unpaired generative adversarial network," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 12, pp. 3374–3383, 2020, doi: 10.1109/JBHI.2020.2999077.

[8] M. M. Islam, F. Karray, R. Alhajj, and J. Zeng, "A review on deep learning techniques for the diagnosis of novel coronavirus (COVID-19)," *IEEE Access*, vol. 9, pp. 30551–30572, 2021, doi: 10.1109/ACCESS.2021.3058537.

[9] J. Chaki and A. Ucar, "An efficient and robust approach using inductive transfer-based ensemble deep neural networks for kidney stone detection," *IEEE Access*, vol. 12, pp. 32894–32910, 2024, doi: 10.1109/ACCESS.2024.3370672.

[10] S. Asif, W. Yi, Q. U. Ain, J. Hou, T. Yi, and J. Si, "Improving effectiveness of different deep transfer learning-based models for detecting brain tumors from MR images," *IEEE Access*, vol. 10, pp. 34716–34730, 2022, doi: 10.1109/ACCESS.2022.3153306.

[11] H. Cui *et al.*, "Automatic segmentation of kidney volume using multi-module hybrid based u-shape in polycystic kidney disease," *IEEE Access*, vol. 11, pp. 58113–58124, 2023, doi: 10.1109/ACCESS.2023.3284029.

[12] A. Prommakhot and J. Srinonchat, "Combining convolutional neural networks for fungi classification," *IEEE Access*, vol. 12, pp. 58021–58030, 2024, doi: 10.1109/ACCESS.2024.3391630.

[13] A. Prommakhot and J. Srinonchat, "VGGNet integration for kidney tumor classification," *12th International Electrical Engineering Congress: Smart Factory and Intelligent Technology for Tomorrow*, 2024, doi: 10.1109/iEECON60677.2024.10537904.

[14] G. Chen *et al.*, "Prediction of chronic kidney disease using adaptive hybridized deep convolutional neural network on the internet of medical things platform," *IEEE Access*, vol. 8, pp. 100497–100508, 2020, doi: 10.1109/ACCESS.2020.2995310.

[15] S. M. M. Elkholy, A. Rezk, and A. A. E. F. Saleh, "Early prediction of chronic kidney disease using deep belief network," *IEEE Access*, vol. 9, pp. 135542–135549, 2021, doi: 10.1109/ACCESS.2021.3114306.

[16] S. D. Pande and R. Agarwal, "Multi-class kidney abnormalities detecting novel system through computed tomography," *IEEE Access*, vol. 12, pp. 21147–21155, 2024, doi: 10.1109/ACCESS.2024.3351181.

[17] S. Lyu, X. Luo, and J. Li, "HyTransMA: a hybrid model using transformer with multi-scale attentions for medical image segmentation," *4th International Conference on Artificial Intelligence, Internet and Digital Economy*, pp. 246–249, 2025, doi: 10.1109/ICAID65275.2025.11034417.

[18] X. Qin, F. M. Bui, H. H. Nguyen, and Z. Han, "Learning from limited and imbalanced medical images with finer synthetic images from GANs," *IEEE Access*, vol. 10, pp. 91663–91677, 2022, doi: 10.1109/ACCESS.2022.3202560.

[19] H. I. Peyal *et al.*, "Plant disease classifier: detection of dual-crop diseases using lightweight 2D CNN architecture," *IEEE Access*, vol. 11, pp. 110627–110643, 2023, doi: 10.1109/ACCESS.2023.3320686.

[20] Y. Qian, H. Tang, Y. Ran, and B. Li, "Target classification in unattended ground sensors with a two-stream convolutional network," *IEEE Sensors Journal*, vol. 23, no. 4, pp. 3747–3755, 2023, doi: 10.1109/JSEN.2022.3226466.

[21] J. N. -Alcazar, S. P. -Castanos, I. M. -Morató, P. Zuccarello, F. J. Ferri, and M. Cobos, "A comparative analysis of residual block alternatives for end-to-end audio classification," *IEEE Access*, vol. 8, pp. 188875–188882, 2020, doi: 10.1109/ACCESS.2020.3031685.

[22] A. K. Rajendran and S. C. Sethuraman, "YogiCombineDeep: enhanced yogic posture classification using combined deep fusion of VGG16 and VGG19 features," *IEEE Access*, vol. 12, pp. 139165–139180, 2024, doi: 10.1109/ACCESS.2024.3414654.

[23] M. N. Islam, M. Hasan, M. K. Hossain, M. G. R. Alam, M. Z. Uddin, and A. Soylu, "Vision transformer and explainable transfer learning models for auto detection of kidney cyst, stone, and tumor from CT-radiography," *Scientific Reports*, vol. 12, no. 1, 2022, doi: 10.1038/s41598-022-15634-4.

[24] H. M. Shahzad, S. M. Bhatti, A. Jaffar, M. Rashid, and S. Akram, "Multi-modal CNN features fusion for emotion recognition: a modified Xception model," *IEEE Access*, vol. 11, pp. 94281–94289, 2023, doi: 10.1109/ACCESS.2023.3310428.

[25] D. Anitha, T. Soujanya, S. Chakraborty, A. Alkhayyat, and R. Revathi, "Oral cancer detection and classification using deep learning with DenseNet121-CatBoost classifier," *2nd IEEE International Conference on Networks, Multimedia and Information Technology*, 2024, doi: 10.1109/NMITCON62075.2024.10698836.

[26] H. M. T. Khushi, T. Masood, A. Jaffar, M. Rashid, and S. Akram, "Improved multiclass brain tumor detection via customized pretrained efficientNetB7 model," *IEEE Access*, vol. 11, pp. 117210–117230, 2023, doi: 10.1109/ACCESS.2023.3325883.

[27] W. Wang *et al.*, "Learning two-stream CNN for multi-modal age-related macular degeneration categorization," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 8, pp. 4111–4122, 2022, doi: 10.1109/JBHI.2022.3171523.

[28] Q. Cai, Z. Gao, J. An, S. Gao, and C. Grebogi, "A graph-temporal fused dual-input convolutional neural network for detecting sleep stages from EEG signals," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 68, no. 2, pp. 777–781, 2021, doi: 10.1109/TCSII.2020.3014514.

[29] C. Ren *et al.*, "Prostate segmentation in MRI using transformer encoder and decoder framework," *IEEE Access*, vol. 11, pp. 101630–101643, 2023, doi: 10.1109/ACCESS.2023.3313420.

[30] S. M. Ganie, P. K. D. Pramanik, and Z. Zhao, "Deep transfer learning for kidney disease detection using CT scan images," *2024 IEEE International Conference on Bioinformatics and Biomedicine*, pp. 5624–5629, 2024, doi: 10.1109/BIBM62325.2024.10822760.
[31] S. Kumar S, K. B, S. J, and M. S, "LMA-CNN: lightweight multi-attention convolution neural network for early diagnosis of chronic kidney disease," *2nd International Conference on Sustainable Computing and Smart Systems*, pp. 1498–1502, 2024, doi: 10.1109/ICSCSS60660.2024.10625428.

BIOGRAPHIES OF AUTHORS

Aekkarat Suksukont received a B.Eng. degree in Electronics and Telecommunication Engineering from Rajamangala University of Technology Thanyaburi (RMUTT), in 2011, and the M.Eng. degrees in Electronics and Telecommunication Engineering from Rajamangala University of Technology Thanyaburi, in 2017. His research interests include speech and image processing, computer vision, and applied AI. He can be contacted at email: aekkarat.s@rmutsb.ac.th.

Anuruk Prommakhot received a B.Eng. degree in Computer Engineering from Rajamangala University of Technology Isan Sakonnakhon Campus in 2015 and an M.Eng. degree in Electronic and Communication Engineering from Rajamangala University of Technology Thanyaburi (RMUTT) in 2018. He completed his Ph.D. degree in Electrical Engineering from RMUTT in 2024. respectively. His research interests the deep learning, optimizers in deep learning, computer vision, and applied AI for medical image. He can be contacted at email: anuruk.p@en.rmutt.ac.th.

Jakkree Srinonchat received a B.Eng. degree in Electrical Engineering from Rajamangala University of Technology Thanyaburi (RMUTT), in 1995 and a Ph.D. degree in Electrical Engineering from the University of Northumbria at Newcastle, UK, in 2005. Since then, he has been a researcher and lecturer Signal Processing Research and Laboratory at RMUTT. His research interests include speech and image processing, robot controllers, machine learning, and applied AI for signal processing. He can be contacted at email: jakkree.s@en.rmutt.ac.th.