INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research

International Journal of Bioinformatics Research and Applications

ISSN online: 1744-5493 - ISSN print: 1744-5485
https://www.inderscience.com/ijbra

Deep learning approach using modified DarkNet-53 for renal cell carcinoma grading

G. Sathish Kumar, G. Uma Maheshwari, C. Selvan, M. Nagasuresh, D. Rasi, P. Swathypriyadharsini, Sathish Kumar Danasegaran

DOI: 10.1504/IJBRA.2025.10064488

Article History:

Received:	17 October 2023
Last revised:	08 February 2024
Accepted:	19 February 2024
Published online:	21 January 2025

Deep learning approach using modified DarkNet-53 for renal cell carcinoma grading

G. Sathish Kumar*

Centre for Computational Imaging and Machine Vision,
Department of Artificial Intelligence and Data Science,
Sri Eshwar College of Engineering,
Coimbatore, Tamil Nadu, India
Email: saathhish@gmail.com
*Corresponding author

G. Uma Maheshwari

Department of Computer Science and Engineering,
Dr. Mahalingam College of Engineering and Technology,
Coimbatore, Tamil Nadu, India
Email: umashanmecse@gmail.com

C. Selvan

School of Computer Science and Engineering,
REVA University,
Bangalore, Karnataka, India
Email: dr.selvan.c@gmail.com

M. Nagasuresh

Department of Information Technology,
Karpagam Institute of Technology,
Coimbatore, Tamil Nadu, India
Email: smnagasuresh@gmail.com

D. Rasi

Department of Computer Science and Engineering,
Sri Krishna College of Engineering and Technology,
Coimbatore, Tamil Nadu, India
Email: priyamudanrasi@gmail.com

P. Swathypriyadharsini

Department of Computer Science and Engineering,
Bannari Amman Institute of Technology,
Erode, Tamil Nadu, India
Email: swa.pspd@gmail.com

Sathish Kumar Danasegaran

Department of Electronics and Communication Engineering,
Sri Eshwar College of Engineering,
Coimbatore, Tamil Nadu, India
Email: dsathish134@gmail.com

Abstract: Accurate and effective diagnostic procedures are required for appropriate treatment planning for renal cell carcinoma, the most frequent form of kidney cancer. Using fusion module, a network dubbed modified DarkNet (MDNet) was developed for image-based small-target detection. We built MDNet on top of a modified version of DarkNet-53, which itself a scale matching approach, to increase its speed and accuracy. By combining the results of several convolutional neural network (CNN) models, the ensemble structure improves classification accuracy. The effectiveness of a classification algorithm using kidney histopathology pictures dataset is measured in accuracy, precision, recall, sensitivity, specificity and F1-score. The results show that the ensemble deep learning method outperforms both standalone CNN models and more conventional machine learning techniques in RCC classification. Overall grade classification accuracy of 98.9%, a sensitivity of 98.2%, and a high classification specificity of 98.7%, in distinguishing tissues.

Keywords: modified DarkNet; convolutional neural network; CNN; ensemble deep learning; kidney cancer; renal cell carcinoma; RCC; whole slide images; WSIs.

Reference to this paper should be made as follows: Kumar, G.S., Maheshwari, G.U., Selvan, C., Nagasuresh, M., Rasi, D., Swathypriyadharsini, P. and Danasegaran, S.K. (2025) 'Deep learning approach using modified DarkNet-53 for renal cell carcinoma grading', *Int. J. Bioinformatics Research and Applications*, Vol. 21, No. 1, pp.1–25.

Biographical notes: G. Sathish Kumar received his BE and ME in Computer Science and Engineering from the Anna University, Chennai, India. He completed his PhD in the Faculty of Information and Communication Engineering from Anna University, Chennai, India. He is currently working as an Assistant Professor in the Department of Artificial Intelligence and Data Science at Sri Eshwar College of Engineering, Coimbatore, India. His areas of interest include data mining, machine learning, information security and bioinformatics.

G. Uma Maheshwari received her BE in Computer Science and Engineering from the Anna University, Tiruchirappalli, India. She completed her ME degree in Computer Science and Engineering from the Anna University, Chennai, India. She is currently pursuing her PhD under the Faculty of Information and Communication Engineering at Anna University, Chennai, India. She is currently working as an Assistant Professor in the Department of Computer Science and Engineering at Dr. Mahalingam College of Engineering and Technology, Pollachi, Coimbatore, India. Her areas of interest include machine learning, bioinformatics, wireless networks and soft computing techniques.

C. Selvan received his BE in CSE in the Manonmaniam Sundaranar University, India in 2002, ME in CSE from Anna University, Chennai, India in 2007 and PhD in CSE from Anna University, Chennai, India in 2013. During his PhD, he was a JRF, SRF in Government College of Technology, Coimbatore under University Grant Commission, New Delhi. He was a PDF in the NIT,

Tiruchirappalli, under UGC, New Delhi, India from June 2017 to June 2022. Currently, he has been working as a Professor in the School of CSE, REVA University, Bangalore since February 2023.

M. Nagasuresh received his Bachelor's and Master's in Computer Science and Engineering from the Anna University, Chennai, Tamil Nadu, India. He worked as an Assistant Professor at the Sri Subramanya College of Engineering and Technology, Palani, Tamil Nadu from 2010 to 2019. After that, he worked as an Assistant Professor at the Professional Group of Institutions, Palladam, Tamil Nadu from 2019 to 2021. He is currently working as an Assistant Professor in the Department of Information Technology at Karpagam Institute of Technology, Coimbatore, Tamil Nadu, India. His research interests include data analytics, machine learning, data mining and deep learning.

D. Rasi pursued her BE in Computer Science and Engineering in 2004 from the Periyar University and ME in Computer Science and Engineering in 2006 from Annamalai University. She was commended with PhD in the Anna University, Chennai during 2017 and currently working as a Professor in Department of Computer Science and Engineering in Sri Krishna College of Engineering and Technology, Coimbatore, Tamil Nadu, India since 2022. She has published 16 international journals, six patents, six books, two book chapters and 20 conferences. Her main research focuses on image processing and data mining. She has 18 years of teaching experience.

P. Swathypriyadharsini is currently working as an Assistant Professor-III in the Department of Computer Science Engineering at Bannari Amman Institute of Technology, Erode, Tamil Nadu, India. She completed her Master of Engineering in Computer Science Engineering (CSE) at the Bannari Amman Institute of Technology, Erode, Tamil Nadu, India and Bachelor of Engineering in CSE at Avinashilingam University, Coimbatore, Tamil Nadu, India. Her research interests include data mining, machine learning, soft computing and artificial intelligence.

Sathish Kumar Danasegaran obtained his BE in ECE from the Kanchi Pallavan Engineering College in Kanchipuram (Anna University, Chennai, India) in 2008 and completed his ME in Communication Systems from Easwari Engineering College in Chennai, India in 2011. He completed his Doctorate degree at the Anna University in 2023. He is working as an Associate Professor in the ECE Department at Sri Eshwar College of Engineering. He has 12 years of academic experience. He is a lifetime member of the ISTE. His research areas include patch antenna design in THz frequency, photonic crystals, photonic bandgap and bioinformatics. He has published 35+ papers in the reputed journals.

1 Introduction

In addition to regulating blood pressure and fluid balance, the kidney plays a role in hormone production. The term 'chronic kidney disease' (CKD) refers to any condition affecting the kidneys' structure or function that lasts longer than three months, according to the Kidney Disease Improving Global Outcomes (KDIGO) definition (Eknoyan et al., 2013). Kidney problems can be caused by a wide variety of medical ailments and disorders (Kovesdy, 2022). Nephrolithiasis, renal cyst formation, kidney stone formation,

tissue from muscles disintegration, haemolytic uremic syndrome, coagulation of the blood, and glomerulonephritis are all examples of diseases that can affect the kidneys (Zhou and Yang, 2020).

Renal cell carcinoma (RCC) is the most common kind of kidney cancer, accounting for almost 90% of all cases. The best treatment plan and prognosis for each patient depend on an accurate assessment of RCC subtypes. The microscopic characteristics of RCC tumours may be better understood with the use of histopathology pictures, which are generated through the analysis of tissue samples. However, pathologists' manual interpretation of these pictures takes a long time, is highly subjective, and might vary from observer to observer. Recently, deep learning (DL) algorithms have established outstanding performance in a number of medical image processing applications. These algorithms can learn discriminative features automatically from enormous datasets and then use those characteristics to create precise predictions. The discipline of digital pathology, which uses histopathology photos to categorise cancer subtypes, has thus paid a lot of attention to them.

Classifying RCC subtypes is difficult due to a number of obstacles, such as a dearth of comprehensive datasets annotated with pinpoint precision. The clear cell subtype accounts for the vast popular of clinical cases, creating a substantial data imbalance; further difficulties include the coherency of RCC cells across subtypes and changes in the appearance of cells within the same subtype at different resolutions. Current RCC categorisation schemes rely heavily on the laborious but invaluable annotation of digital pathology slides. Non-medical and medical image analyses, such as facial recognition and x-ray analysis, rely heavily on artificial intelligence (AI) (Jang and Cho, 2019). The 'neurons' at the heart of artificial neural networks collect data from numerous sources, which are then merged mathematically using nonlinear activation functions. Using artificial neural networks, DL is an area of AI that involves stacking numerous layers of these units. The complex hidden layer topologies of DL techniques like convolutional neural networks (CNNs) make them ideal for image processing. CNNs have outperformed human experts in various clinical applications (Rajkomar et al., 2018).

When it comes to tasks like visualisation and computer vision, CNNs stand out as a sort of DL neural network that excels. CNNs have found widespread use in radiology, and more recently, oncological surgical pathology (Kather et al., 2020; Bulten et al., 2020), both of which rely heavily on image analysis to provide accurate diagnoses. Using cancer histopathology photos, CNNs may be trained directly on the unprocessed information that does not require human interpretation of intermediary stages (Pantanowitz et al., 2020), allowing them to recognise molecular subtypes of tumours. Computerised semantic segmentation of histology into multiple histological compartments is the primary focus of the few current uses of DL in transplant histopathology and nephrology (Echle et al., 2020; Hermsen et al., 2019; Jayapandian et al., 2021). Furthermore, molecular pattern study of renal allograft rejection and allograft loss prediction systems (Bouteldja et al., 2021) have been created. Nevertheless, no comprehensive DL biomarkers for transplant histopathology are currently available.

To classify RCC subtypes from kidney histopathology pictures, we present an ensemble DL approach in this work. Using an ensemble method, numerous DL models are combined to improve the classification system's overall efficiency and reliability. The main contribution of this work is as follows:

1 Increase accuracy and decrease the possibility of misclassification by combining the predictions of many models.

2 The histopathological pictures are pre-processed in our manner to increase feature visibility and decrease noise.

3 In order to detect tiny targets in images, developed a network called modified DarkNet (MDNet) that makes use of a fusion module to provide more background information about the targets.

4 Present a scale-matching technique to help choose between several scales for detecting tiny targets. To improve the speed and accuracy of MDNet, we used a modified version of DarkNet-53 as the network's foundation. It is based on the scale matching technique.

A huge annotated dataset of kidney histopathology pictures containing several RCC subtypes is used to train the CNNs. We use numerous quantitative indicators, including as accuracy, precision, recall, and F1-score, to assess the efficacy of our ensemble DL method. To further evaluate the efficacy of the ensemble strategy, we also compare our results with those derived from individual DL models.

2 Related works

Here, we provide a synopsis of the work in the fields of semantic segmentation, neural network segmentation and segmentation of nuclear structures. Mathematical morphological processes similar as those presented in Faridi et al. (2016) are commonly used in conventional approaches to nucleus segmentation from H&E-stained histopathology pictures. Before the development of DL, supervised-learning-based segmentation techniques depended heavily on manually built features that classified pixels separately. Commonly, a patch is taken from the data and put into a classifier like random forest. Despite their ease of use, they were limited in their segmentation of different types of tissue pictures, which is typical in the area of pathology.

Cireşan et al. (2012) proposed the use of a CNN to tackle the difficult task of segmenting brain tissue from electron microscopy pictures. This model proposed employing a shallow CNN to make a binary foreground/background determination from picture patches in a single pass. That is to say, each patch in the picture must be processed many times using forward passes, each of which labels a single pixel as foreground or background. The large number of processes required to segment a complete image is a major drawback of this method. Before employing a DNN to segment patches, Pan et al. (2017) recommend doing a sparse reconstruction and a morphological post-processing phase. Jung et al. (2019) proposed a model for colour normalisation based on deep convolutions using a Gaussian mixture of distributions, which successfully clusters pixels without ignoring the nucleus structure.

An internally dataset of only 26 kidney biopsies was used to train a CNN. On the other hand, a two-stage DL technique was realised and evaluated on 47 whole slide images (WSIs) in Bueno et al. (2020). DL algorithms are applied on lung ultra sound images to detect and to classify the COVID-19 (Babu et al., 2023). This approach included first segmenting the patterns in the images and then categorising the segments

using an AlexNet-like framework. Additionally, in Temerinac-Ott et al. (2017), many staining methods for predicting sick and non-diseased kidney tissues were examined within the context of a classification model. Deciphering the DL techniques for brain image analysis along with CNN is applied to predict the Alzheimer's disease (Mohapatra et al., 2023). The research, which included samples from six patients, shows that DL methods are more effective than humans at this task. Good glomerulus identification accuracy was achieved using an AlexNet-based sliding window technique applied to the WSIs in Gallego et al. (2018). Additionally, in Gu et al. (2022), the semantic segmentation of glomerulus regions was accomplished with the use of an ensemble classifier. Homomorphic encryption along with the statistical approach is used to hide the sensitive data from the third parities (Kumar et al., 2023). In the scenario of mice-glomeruli segmentation/detections, similar methodologies may be discovered (Bukowy et al., 2018; Stritt et al., 2020), but with a somewhat greater picture count. Indeed, the lack of unified standards to be employed for the segmentation of glomeruli is one of the key challenges in the case of human kidneys.

Histopathological pictures were employed in a four-class renal neoplasm distinction using morphological, textural and wavelet-based approaches (Kocak et al., 2018; Asteris et al., 2022). Grey-level co-occurrence matrix (GLCM) basic characteristics including contrast, correlation, energy, homogeneity, and entropy fed the Bayesian classifier to an acceptable 90% accuracy. Computed tomography texture analysis (CTTA) (Raman et al., 2014; Shahin et al., 2023) is another area where textural characteristics have proven useful. Simple statistical aspects of raw image pixel intensities, including average intensity, entropy, or skewness, were extracted as biomarkers for tumour distinction. Due to the greater incidence rate of this disease, we were motivated to build larger datasets, making DL approaches ideal for use in lung cancer detection tasks. Autoencoders (Polat and Homay, 2019), 3D-CNNs (Alakwaa et al., 2017), and deep belief networks have all been shown to reach state-of-the-art results in this field.

Most DL cancer detection methods use slide-level annotations. These systems use poorly supervised learning to overcome the absence of big datasets with precisely localised annotations. Multiple-instance learning (MIL) involves a patch-level CNN and slide-level algorithm. The authors trained neural networks on full WSIs for lung cancer type categorisation in Chen et al. (2020). They trained traditional CNNs with large picture inputs using the unified memory (UM) mechanism and GPU memory optimisation approaches without changing training pipelines or model designs. Adenocarcinoma and squamous cell carcinoma classification area under the curves (AUCs) were 0.9594 and 0.9414, respectively, exceeding MIL methods. These articles used Resnet and Inception-V3 CNN architectures. Pan-cancer whole-slide picture categorisation was simplified with PathCNN (Bilaloglu et al., 2019). PathCNN training data combines lung, kidney, and breast cancer tissue WSIs. The suggested architecture converged quicker and used less memory than Inception-V3. In Gao et al. (2020), researchers used \minimal-point-based annotation to identify RCC areas semi-supervisedly. Their RCC subtyping classification approach uses a mixed loss method. Although it requires partially annotated data, the minimal-point-based classification model beats whole-slide models by 12%. In computational pathology, attention-based models performed similarly to MIL methods. We propose a multiscale MIL architecture to improve traditional MIL classification performance. By weighting each identified feature, attention processes can improve interpretability and visualisation (Chaudhari et al., 2021). DL techniques, especially CNNs, are commonly used for

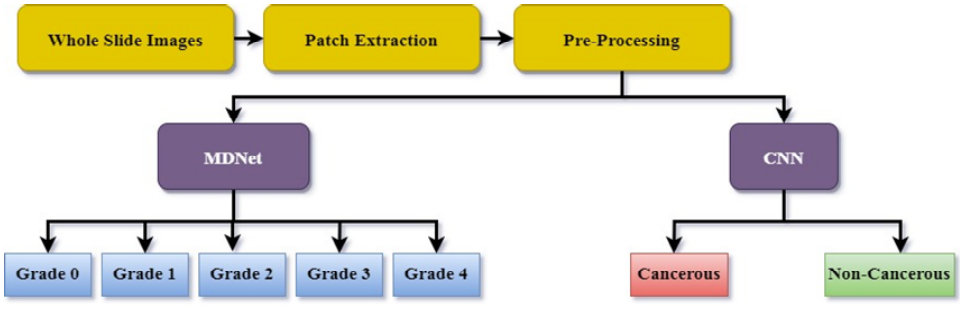
processing the semantic segmentation. U-Net, fully convolutional network (FCN), and DeepLab are employed for semantic segmentation. Neural network segmentation uses neural networks for segmenting images or data. This can include semantic segmentation and instance segmentation (Magneta et al., 2023). Lu et al. (2021) introduced clustering-constrained attention multiple-instance learning (CLAM). Fuzzy ensemble techniques like majority voting and Choquet fuzzy are pooled to find the decision marks for the transfer learning model to detect the COVID-19 patients. This method achieved 99.57% of accuracy (Gavriilaki et al., 2021; Mahanty et al., 2021).

Tsuneki and Kanavati (2021) refined the pre-trained model using a partial transfer learning approach. Using their technique, they have attempted to categorise the extremely uncommon poorly differentiated colorectal adenocarcinoma. Their method comprises tweaking a limited subset of the pre-trained model's batch normalisation (BN) layer settings. They employed the ImageNet-trained EfficientNetB1 model. The model's ROC AUC was 0.94. Instead of utilising WSIs, Byeon et al. (2022) have classified colorectal lesions into six different types using digital pictures of pathology slides. Using DenseNet-161, they averaged 97.3% accuracy while using EfficientNet-B7, they averaged 95.9% (Chanchal et al., 2023). Classification algorithms are made combined with refined diagnostic model and the classifier based on multilayer perceptron for diagnosing the diabetic retinopathy. This model archives 70.7% of accuracy and 75.8% of feature selection capacity (Odeh et al., 2022).

3 Materials and methods

In this paper, we present an ensemble DL technique that combines numerous sub-networks of varying architectures. The ensemble model makes the classification call by combining the results of numerous simpler systems. This method draws on the synergistic benefits of many DL architectures to improve the model's efficiency and scalability. With the use of transfer learning and data augmentation, among other methods, the ensemble DL model will be trained on the labelled dataset to boost generalisation and prevent overfitting. Through cross-validation and Bayesian optimisation strategy, the model's hyperparameters will be adjusted.

Figure 1 Overview of the proposed methodology (see online version for colours)

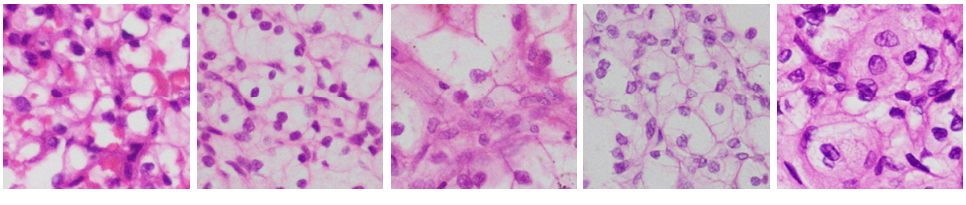


Using pixel thresholding, patches were generated from the photos with a 50% overlap and the background was eliminated. The patches from non-cancerous and malignant slides were put into the deep network, and the network determined whether the patches were cancerous or not. Two distinct DL classifiers were used to analyse the patches. One for the malignant and non-cancerous categories in the binary categorisation system. The photos of non-cancerous conditions that were given into model 2 in order to grade and classify them. The proposed model overview is shown in Figure 1. Patches were recovered from photos and the background was eliminated using pixel thresholding. Patches from normal and malignant slides were put into the deep network, and they were classified as cancerous or non-cancerous.

3.1 Dataset

Figure 2 from the KMC kidney histopathology material set, which was advanced by Chaudhari et al. (2021), validates the occurrence of both benign (Grade 0) and malignant (Grades 1–4) renal clear cell carcinomas. The dataset is partitioned into two parts: the training set and the test set. Only around 20% of the total photos were used for testing, whereas 80% were used for training. Cropping non-overlapping parts of training patches yielded the validation set.

Figure 2 (a) Grade 0 – a normal glomerulus structure (b) Grade 1 – morphology is very similar to normal nuclei (c) Grade 2 – irregular contour (d) Grade 3 – clearly visible tumours (e) Grade 4 – contains tumour giant cells (see online version for colours)

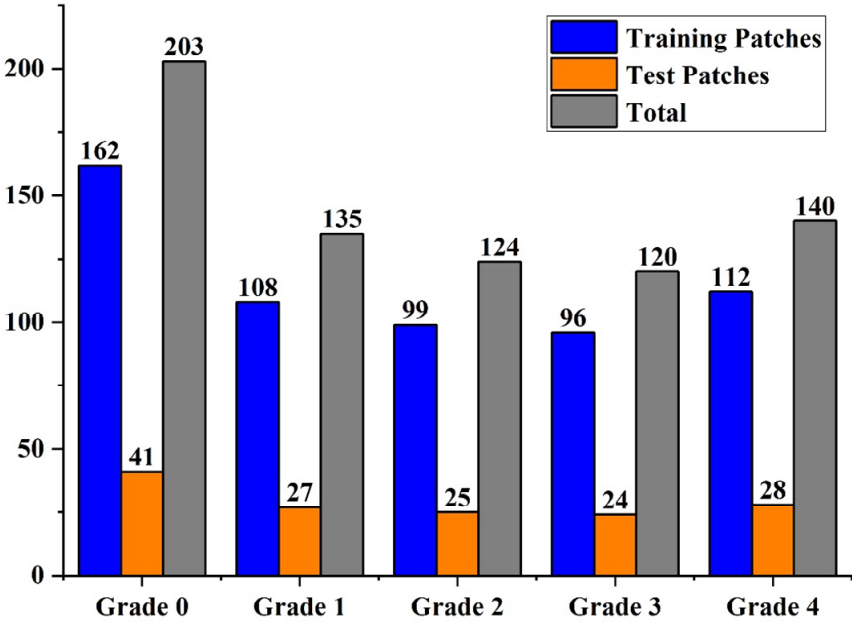


(a) (b) (c) (d) (e)

To further boost the diversity of the training set, horizontal and vertical flipping were utilised as data augmentation approaches. The sample images are shown in Figure 2. The training set consisted of 3,442 patches from five different classes after random cropping and data enhancement. We have not used test set data augmentation and have instead evaluated our models using the raw data from the test set. Figure 3 displays the dataset's dispersion.

Splitting datasets into training and testing sets is a crucial step in machine learning to evaluate model performance on unseen data. Most commonly, the researchers will randomly divide the dataset into two parts, one for training the model and the other for testing its performance. In this research work, train-test split ratio approach is employed to decide on the ratio of data to be used for training and testing. A split of 80%–20% is considered for the experimentation, where the majority of the data is used for training and the rest for testing.

Figure 3 Data distribution of the dataset (see online version for colours)



Note: 693 of these spots are classified as normal, 648, 708 as Grade 1 cancer, 735 as Grade 3 cancer, and 648 as Grade 4 kidney cancer.

Source: Chaudhari et al. (2021)

3.2 Pre-processing

Unfortunately, a DL model cannot comprehend the picture slides due to their sheer size. To maximise an algorithm for DL and get high accuracy, it is necessary to generate a sufficient number of examples; this is achieved by constructing image slides using patches of photos. The dataset was used to create overlapping patches with the following dimensions: small = 250 × 250, medium = 350 × 350 and big = 450 × 450. Each image slide now has its own folder, within of which are subfolders for the three different patch sizes. Each patch was spaced out so that there was a 50% overlap with the next. Because of the overlapping patches, the DL framework was able to acquire different perspectives inside the tissue. Background-dominated areas were eliminated, leaving around 70% for training and the remaining 30% for testing.

By standardising and enriching data, pre-processing boosts the precision of the suggested approach. Image data augmentation is used to advance the model's capability to learn and generalise. The term 'dataset augmentation' refers to a method of artificially expanding a training dataset by the addition of new, up-to-date photos. Initiating the data I/O interface allows for the selection of the data augmentation class. To reduce the overfitting error rate of a model, data augmentation is a useful method. Among the metrics employed in augmentation is shown in Table 1.

Table 1 Metrics for data augmentation

Pre-processing technique	*Description*	*Value*
Normalisation	Scaling pixel values to a specific range	0.2
Resizing	Adjusting image dimensions	256 × 256
Cropping	Removing irrelevant image borders or regions	224 × 224
Rotation	Rotating the image for better alignment	90 degrees
Gaussian blur	Smoothing the image to reduce noise	Kernel size: 3 × 3
Contrast adjustment	Enhancing image contrast for better visibility	Gamma: 1.2
Colour augmentation	Modifying colour channels for data variation	Brightness: 0.2
Noise reduction	Removing noise artefacts from the image	Standard deviation: 1.5

3.3 Modified DarkNet-53

Although DCNN-based target detectors have shown impressive performance, these detectors are only ever used for broad classes of targets. There is little research focused on detecting tiny targets. In order to improve the identification of tiny targets, the researcher up-samples the deep feature maps using deconvolution and fuses them with the shallow ones. Due to its efficiency and ease of implementation, we use DarkNet-53 as our primary backbone design as shown in Figure 4. DarkNet-53 borrows from the concept of residual networks and uses it to generate a series of 3 × 3 convolutions followed by 1 × 1 convolutions.

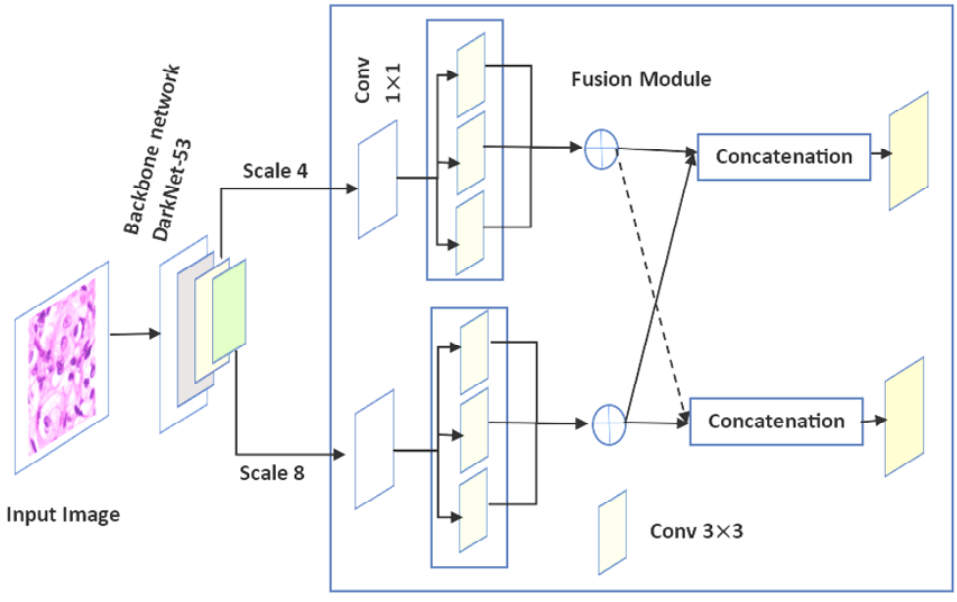
DarkNet-53 has three different scales, with corresponding down sampling of 8, 16, and 32 for its feature maps. The CNN DarkNet-53 has 53 layers. The YOLOv3 object identification mechanism relies on this. By merging Resnet's features, it solves the gradient issue that arises from a structure that is excessively deep and ensures that all characteristics are represented optimally. There are several layers of convolution 1 × 1 and 3 × 3 and residual blocks. The following is an explanation of the convolutional layer:

$$c_x^y = \sum_{n \in I_m} c_m^{y-1} * J_{mx}^y + K_x^y \tag{1}$$

In equation (1), the input picture is convolved by multiple convolution kernels to create m exclusive feature maps c_x^y, with the y^{th} feature map helping as the depiction for layer y. The convolution operation is epitomised by the symbol $*$. I_m stands for the image's feature vector, while J_{mx}^y stands for the y^{th} component of the y^{th} layer's convolution kernel. The BN layer follows, which is crucial.

$$c_o = \frac{\vartheta\left(c_x^y - \partial\right)}{\sqrt{\omega^2 + \varphi}} + \gamma \tag{2}$$

Figure 4 MDNet-53 architecture (see online version for colours)



Notes: Scales 4 and 8 are employed in a scale matching method for detecting tiny things. The fusion module is utilised to improve the surrounding context data of the tiny targets.

The persistent offset is indicated by ϑ, the average of the outputs is symbolised by ∂, the variance of the inputs is denoted by ω, and the result of the convolution computation is indicated by c_o in equation (2). c_o stands for the output of BN. BN is applied to the output to ensure that the coefficient distribution of the same batch of eigenvalues is maintained. The next layer is a convolutional one, which helps speed up network convergence and prevents over-fitting. The layer that follows is commonly referred to as the activation layer. The activation function of DarNet53 consists of a leaky rectified linear unit (ReLu) layer.

$$I_y = \begin{cases} m_y, & \text{if } c_o \geq 0 \\ \dfrac{m_y}{d_y}, & \text{if } c_o < 0 \end{cases} \tag{3}$$

In equation (3), m_y stands for the input value, xj for the activation value, and d_y for the fixed parameter in the interval [–1, +∞]. The pooling layer is also a crucial part of this network. The network's weights are downsampled at this layer. In this network, the max-pooling layer is employed. The last illustration shows how features (weights) may be integrated into a single layer. In the output layer, these properties are categorised for final processing. This model has a depth of 53, a size of 155 MB, 41.6 million parameters, and an input picture size of 256 × 256 pixels. The layer-by-layer architecture is described in Table 2.

Table 2 The layered architecture of MDNet for grading of the kidney cancer

Layer	*Output shape*	*Number of parameters*
Input	(224, 224, 3)	-
Conv3×3	(112, 112, 32)	896
Max pooling	(56, 56, 32)	-
Conv3×3	(56, 56, 64)	18,496
Max pooling	(28, 28, 64)	-
Conv1×1	(28, 28, 32)	2,080
Conv3×3	(28, 28, 128)	36,992
Conv1×1	(28, 28, 32)	4,128
Conv3×3	(28, 28, 128)	36,992
Residual	(28, 28, 128)	-
Conv1×1	(28, 28, 64)	8,256
Conv3×3	(28, 28, 256)	147,712
Conv1×1	(28, 28, 64)	16,448
Conv3×3	(28, 28, 256)	147,712
Residual	(28, 28, 256)	-
Conv1×1	(28, 28, 128)	33,024
Conv3×3	(28, 28, 512)	590,336
Conv1×1	(28, 28, 128)	65,664
Conv3×3	(28, 28, 512)	590,336
Residual	(28, 28, 512)	-
Conv1×1	(14, 14, 256)	131,328
Conv3×3	(14, 14, 1,024)	2,360,832
Conv1×1	(14, 14, 256)	262,400
Conv3×3	(14, 14, 1,024)	2,360,832
Residual	(14, 14, 1,024)	-
Conv1×1	(14, 14, 512)	524,800
Conv3×3	(14, 14, 2,048)	9,438,208
Conv1×1	(14, 14, 512)	1,049,088
Conv3×3	(14, 14, 2,048)	9,438,208
Residual	(14, 14, 2,048)	-
Avg. pooling	(1, 1, 2,048)	-
Flatten	(2,048)	-

We begin by applying a 1 × 1 convolution on the features extracted from MDNet-53 in order to minimise their dimension. After that, we dilate the receptive field by 1, 2, and 4 times using three 3 × 3 convolutions. By fusing the data from many receptive fields, we are able to better identify tiny objects. Various characteristics' fields of reception have unequal impacts on the final product. As a result, we suggest enhancing features with varying receptive fields with learnable weights. Let us pretend that the weight of each

receptive field is W_x and that its characteristics are f_x. After that, you may determine the combined characteristics by,

$$f_{fuse} = \sum_{x=0}^{k} W_x \cdot f_x \tag{4}$$

The following formula is used to generate values for the learnable weight between 0 and 1.

$$W_y^* = \frac{W_y}{\sum_{x=0}^{k} W_x} \tag{5}$$

Training the same parameters with different receptive fields increases the network's robustness.

3.3.1 Loss function (GIOU)

The location of the minor tumours is regressed using generalised intersection over union loss (GIOU loss). GIOU can be calculated with the help of the equation that follows:

$$G_{T_p,T_a} = \frac{\left|T_p \cap T_a\right|}{\left|T_p \cup T_a\right|} - \frac{\left|\frac{T}{T_p \cup T_a}\right|}{|T|} \tag{6}$$

where T_p is the real target and T_a is the box that was projected. The smallest convex area encompassing both T_p and T_a is denoted by T. The GIOU loss may then be determined by,

$$L_G = 1 - G_{T_p,T_a} \tag{7}$$

Focal loss is used to regress the confidence of the tiny pixels so that we may focus more on difficult samples and less on the negative or positive ones.

$$L_c = \begin{cases} -\alpha\left(1-c_p\right)^{\gamma} \times \log c_p, & T_p = 1 \\ -(1-\alpha)c_p^{\gamma} \times \log\left(1-c_p\right), & T_p = 0 \end{cases} \tag{8}$$

where c_a is the confidence in the ground truth and c_p is the confidence in the prediction. The values for α and γ are defined as 0.25 and 2. Binary cross-entropy reduction is used to regress the class of low-enough-enough objectives.

$$L_n = c_a \log c_p - \left(1-c_a\right)\log\left(1-c_p\right) \tag{9}$$

c_a is the predicted class and c_p is the predicted class, n is the number of classes. Then, the total loss of MDNet is calculated as,

$$L_{total} = \sum_{x=0}^{k} L_G^s + L_c^s + L_n^s \tag{10}$$

where s is the scaling of the model proposed in this paper.

3.4 Convolutional neural network

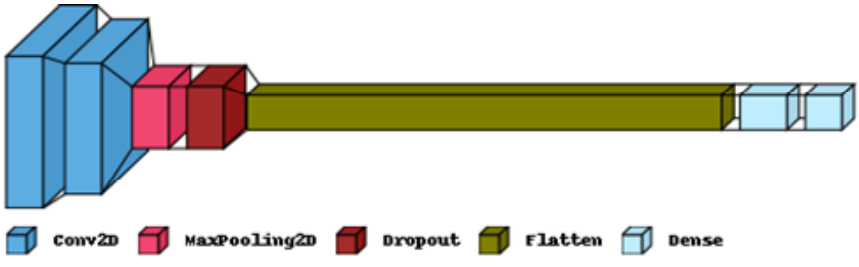
All the photos in the collection are scaled down at the outset. Since most photos in the collection are larger than 50 × 50 × 3, this is necessary. Images are 'standardised' by dividing them by 255, which ensures that all pixel values fall within the range of 0 to 1. This is done to speed up the model-training process and improve the model's training efficiency. It safeguards us from the dreaded 'exploding gradients' issue as well.

- To fix the data imbalance, we under sample the majority group at random. Minority class samples are increased while majority class samples are decreased.
- To improve the model's generalisability, data augmentation is used to lessen the network's propensity to overfit the training set.

Techniques including rotation, horizontal flipping, and vertical flipping are used to guarantee consistency. In order for the network to accurately predict a wide variety of classes, rather than just the ones on which it was trained, these measures are essential.

Convolutional layers, pooling layers, and fully linked layers make up a CNN's basic architecture (see Figure 5). In the bottom layer, connected neurons' outputs are computed. A dot product of the weights and the area is used to determine each. Common filter sizes for input images are 3 × 3, 5 × 5 or 8 × 8 squares. These filters learn the patterns that appear everywhere in a picture and then scan the image using a sliding window. The stride refers to the space between each filter. If the stride hyperparameter is lower than the filter dimension, the convolution is expanded to overlap windows.

Figure 5 Basic architecture of CNN for the detection of kidney cancer (see online version for colours)



In the convolutional layer, the image is compressed with a series of filters one at a time.

$$f_{x,y} = \sum_{m=1}^{l}\sum_{n=1}^{l} w_{m,n} \cdot I_{x+m-1,y+n-1} \tag{11}$$

$f_{x,y}$ is the convolutional output, the kernal size is indicated by l, I_x is the input data. $w_{m,n}$ is the weight. The quantity of processing and network parameters can be lowered by using a pooling layer to gradually shrink the size of the depiction. Maxpooling is the method applied in this case.

$$f_{x,y} = \max\left\{I_{x+m-1,y+n-1} \;\forall\, 1 \le m \text{ and } 1 \le n \le l\right\} \tag{12}$$

The activation function used here are,

$$ReLu = \begin{cases} I, & \text{if } I > 0 \\ 0, & \text{if } I \le 0 \end{cases} \tag{13}$$

$$sigmoid(s) = \frac{1}{(1+e^{-aI})} \tag{14}$$

This layer receives a volume from the previous layers as input and generates a vector of length n, where n is the number of classes. To flatten a 3D feature map into a 1D feature vector, we utilise the flatten layer tool. The model was trained with a batch size of 256 and 300 epochs. Adam optimiser, with an initial learning rate of 0.00001, is used to do weight updates and minimise loss.

4 Experimental setup

Python's PyTorch DL framework has been used to realise the suggested and current DL architectures. We have used an NVIDIA Tesla K80 GPU for training and inference to speed up the development process. We checked that the dataset was free of bad information and labels before beginning the training process. Our dataset was segmented into a training set, a validation set and a test set. We used dynamic data augmentation and created ground truth distance maps during training. We train the model with a specified set of hyperparameters and then assess its accuracy using the validation set. We run the identical procedure many times with different hyper-parameter values and pick the one that improves the F1-score the most.

To facilitate simultaneous training and validation on the computer, a batch size of 64 was used. Adam optimiser was used with the default settings of $\alpha = 0.01$ for the initial learning rate, $\beta 1 = 0.9$ for the moment decay rate, and $\beta 2 = 0.999$ for the final decay rate. Categorical cross-entropy is used to derive the loss function for a five-class categorisation task. The models are constructed and trained for 300 epochs in this configuration.

4.1 Results and discussion

The sample data validates the accuracy of both the MDNet model and the CNN for detecting malignant and non-cancerous kidney tumours. The model's predictive performance is rather good, since it is able to determine the proper grades in a large percentage of tests. However, misclassifications do happen, highlighting the difficulties in categorising kidney cancer. The model's efficacy was measured using the confusion matrix, precision, recall, F1-score and weighted-average. In addition, data pre-treatment procedures, such as data cleansing, dimensionality reduction (feature reduction), and data normalisation, are employed. Additionally, the following step partitions the raw data into groups. The training set is used to teach the DL classification algorithm how to properly categorise data. The test set is used to further assess the recital of the trained classification model. Evaluation of the data can reveal the model's precision. The occurrence of cancer is on the rise. The non-appearance of symptoms in the early stages of the disease makes diagnosis challenging. Early diagnosis increases the possibility of an individual's recuperation and ultimate healing. Cancer is notoriously hard to identify early and often returns after being treated. The process of categorising cancers is critical.

The results of each cross-validation set were weighted and averaged to find an overall performance, which works well with unbalanced datasets. The below listed formulas can be used to calculate various performance indicators.

$$Accuracy\left(A_{cy}\right)=\frac{Tr_p+Tr_n}{Tr_p+Tr_n+Fa_p+Fa_n} \tag{15}$$

$$Sensitivity\ (SE)=\frac{Tr_p}{Tr_p+Fa_n} \tag{16}$$

$$Specificity\ (SP)=\frac{Tr_n}{Tr_n+Fa_p} \tag{17}$$

The precision of the proposed model is revealed in Figure 6. Figures 6(a) and 6(b) display the accuracy of the MDNet and CNN models, correspondingly. Figures 6(c) and 6(d) depict the MDNet and CNN model losses, respectively. As can be exposed in Figure 6, the MDNet outperforms competing models.

In spite of the limited number of samples, the experiment revealed that MDNet models performed the best when it came to correctly classifying the validation set with an accuracy of up to 98.9% when grading kidney cancer.

The CNN model uses a dropout regularisation of 50% to balance for overfitting and features two convolution layers with 32 and 64 kernels. Next, a dense layer is further to the vectorised picture in the form of a flattened layer. In all levels except the output layer, which employs the Softmax activation function, the ReLu is hired as the activation function. This model has been trained for 300 iterations with a 128-person training batch. Both the training and validation losses are 0.15. There is not much of a distinction between the model's training set and validation set results.

$$Precision\left(PR_{re}\right)=\frac{Tr_p}{Tr_p+Fa_p} \tag{18}$$

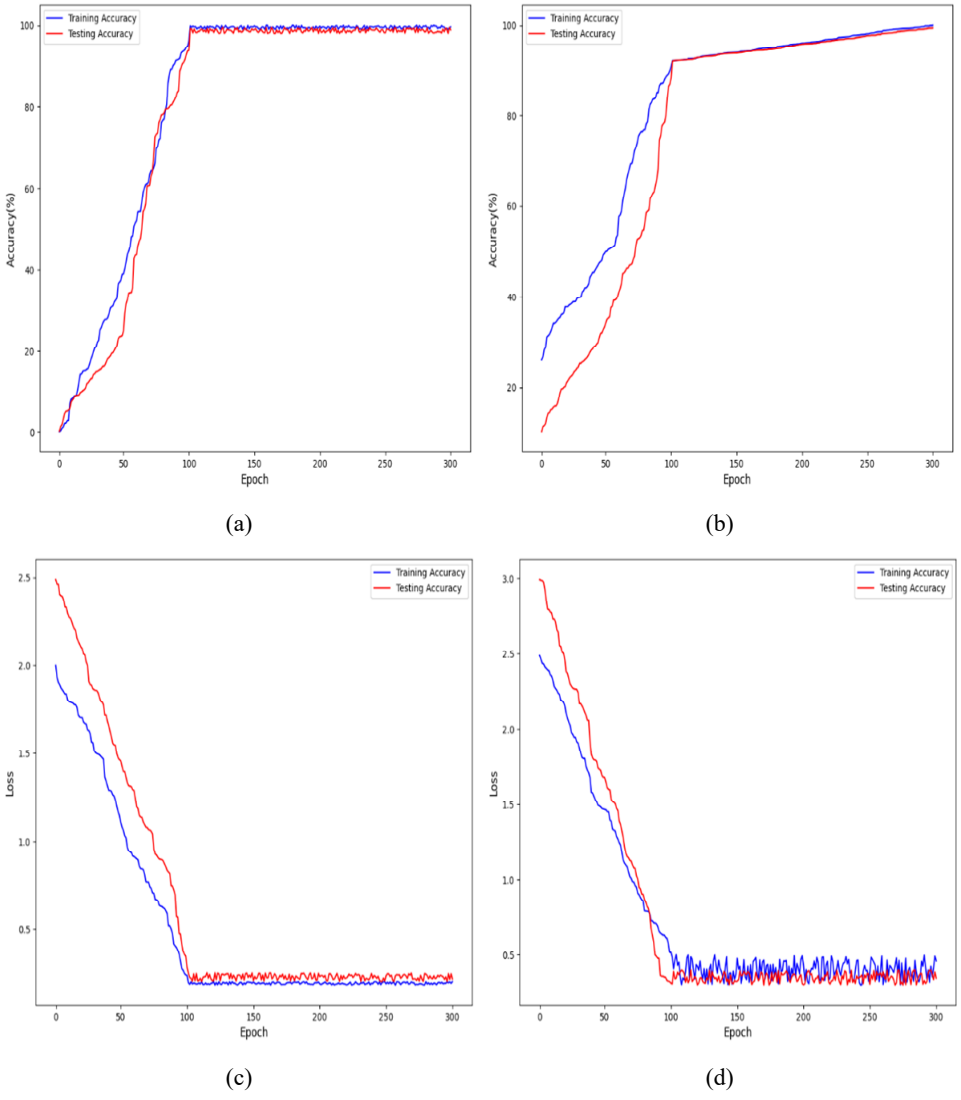
$$Recall\left(RE_{cl}\right)=\frac{Tr_p}{Tr_p+Fa_n} \tag{19}$$

$$F1\text{-}score\left(F1_s\right)=2\times\frac{PR_r*RE_{cl}}{PR_r+RE_{cl}} \tag{20}$$

The efficiency of a categorisation system may be restrained with a table termed a confusion matrix. It summarises the results of a categorisation task, both anticipated and real. A confusion matrix can be used to evaluate a model's ability to forecast whether a patient has cancer or not in the setting of cancer detection. There are four possible values in a sample confusion matrix for cancer detection: true positive (Tr_p), false positive (Fa_p), true negative (Tr_n) and false negative (Fa_n). Figure 7 demonstrates the confusion matrix of the proposed MDNet and CNN.

The greatest F1-scores (98% to 99%) were all achieved by MDNet, which has a considerably deeper layer than the other baseline models. Tables 3 and 4 show the recital assessment of the proposed models. MDNet and CNN performed slightly great on all three datasets. This demonstrates that one may improve prediction accuracy by using an MDNet with a deeper layer.

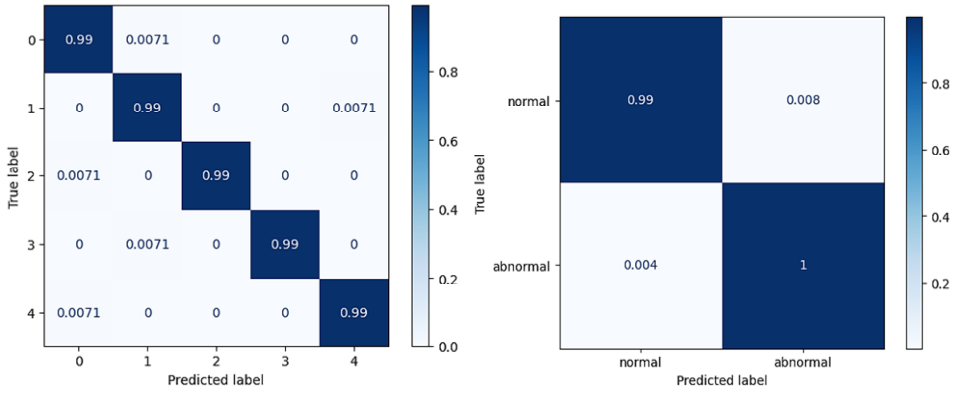
Figure 6 Comparing the MDNet model's accuracy and loss against those of a CNN model on the dataset, (a) MDNet model reliability (b) CNN model accuracy (c) the loss of the model MDNet (d) the loss of CNN model (see online version for colours)



Specificity evaluates how many real negatives the grading system accurately identifies. It shows the system's RCC/non-RCC discrimination. Grade 0 has a specificity of 0.98, meaning 98% of non-Grade 0 RCC instances were accurately recognised as such. Precision, also known as positive predictive value, is the percentage of real positive instances in the grading system's positive cases. It verifies the system's optimistic predictions. For Grade 0, the accuracy is 0.99, implying that 99% of Grade 0 instances were RCC. As previously indicated, recall is sensitivity. It measures how many positive examples the grading system successfully detected. Grade 0 RCC has a recall of 0.99, suggesting 99% of cases were accurately diagnosed as Grade 0. Precision and recall form

the F1-score. It gives accuracy and recall equal weight in one metric. When there is a positive-negative imbalance, it is utilised. For Grade 0, the F1-score is 0.98, suggesting balanced precision and recall.

Figure 7 The confusion matrix of the MDNet and CNN (see online version for colours)



Note: The MDNet performs better than the single CNN model.

Table 3 The performance metrics on the model MDNet

	Sensitivity	*Specificity*	*Precision*	*Recall*	*F1-score*
Grade 0	0.99	0.98	0.99	0.99	0.98
Grade 1	0.98	0.98	0.99	0.98	0.99
Grade 2	0.98	0.99	0.99	0.98	0.99
Grade 3	0.99	0.99	0.99	0.99	0.98
Grade 4	0.99	0.98	0.99	0.99	0.99
Accuracy				98.9	
Macro avg.			0.99	0.99	0.99
Weighted avg.			0.99	0.99	0.99

ROC curves show classification model performance. It is often used to compare classification methods or models. The ROC curve compares TPR and FPR at different categorisation levels. You appear to have a model that predicts RCC Grades 0–4. AUC summarises model performance. Higher AUCs suggest better RCC grade differentiation. These AUC values show that the model distinguishes RCC grades well.

Table 4 The recital metrics on the model CNN

	Sensitivity	*Specificity*	*Precision*	*Recall*	*F1-score*
0	0.99	1.00	1.00	0.99	0.98
1	0.99	0.99	1.00	0.99	0.99
Accuracy			99.89		
Macro avg.	1.00	0.99	1.00	1.00	0.99
Weighted avg.	0.99	1.00	1.00	0.99	0.99

Figure 8 The ROC-AUC curve of the proposed model on grade classification (see online version for colours)

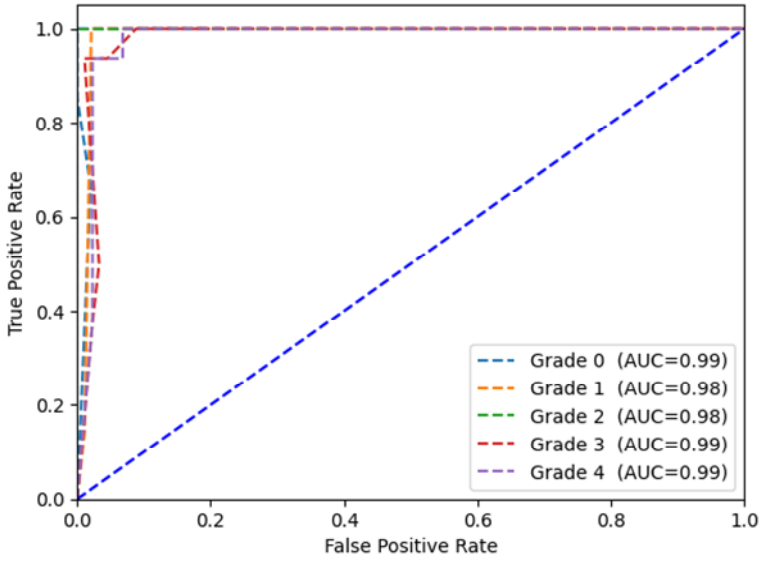
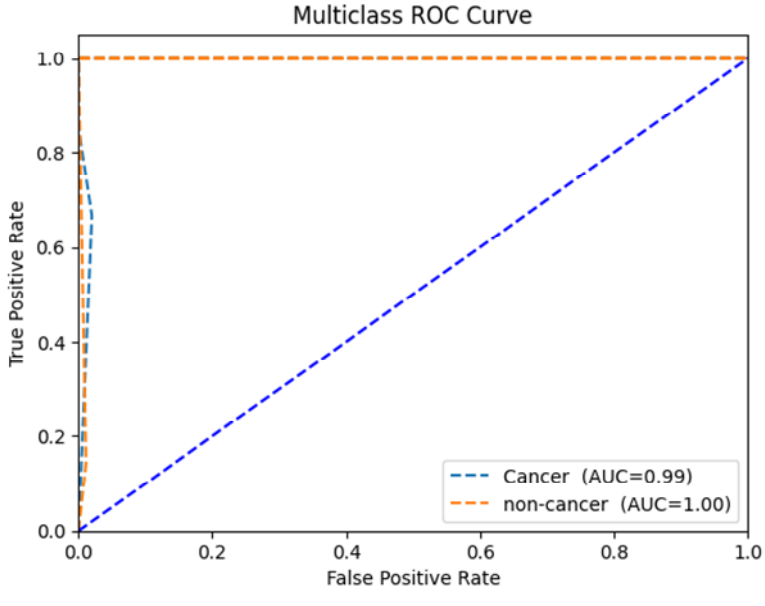


Figure 9 The ROC-AUC curve of the proposed model on cancer and non-cancer classification (see online version for colours)



1.0 is a perfect classifier, whereas 0.5 is a random classifier. Your model classifies RCC grades accurately with AUC values from 0.98 to 0.99. The RCC classification model's ROC curve for these AUC values is TPR vs. FPR as shown in Figure 8. The ROC curve's

shape and properties can reveal the model's appropriate classification threshold and sensitivity-specificity trade-off.

The AUC values, AUC = 0.99 and 1.00 for both cancer and non-cancer, indicate that the classification model has a very high discriminatory power for both classes as shown in Figure 9. Since both the cancer and non-cancer AUC values are 0.99 and 1.00, it suggests that the model has very high sensitivity (true positive rate) and specificity (true negative rate) for both classes. This indicates that the model can effectively distinguish between cancer and non-cancer instances. On the ROC curve, the higher the curve is to the top-left corner, the better the model's performance. With an AUC of 0.99 and 1.00 for both classes, the ROC curve for your classification model would be close to the top-left corner, indicating excellent performance.

5 Ablation study

To see how data augmentation affects a model's training performance, we conducted this experiment with the first cross-validation set of the mixed dataset. After 300 epochs, the ultimate accuracy and loss on both the training and validation set are shown in Table 5. Table 4 shows that overfitting has happened without any additional data since the model accomplishes well on the training set but poorly on the validation set. The MDNet model in particular saw a boost in validation accuracy from 98.9% to 99.89% after receiving more data. As the model is learning freshly converted pictures, the training accuracy of other models will naturally decrease during data augmentation. As can be seen in Table 5, the validation loss has decreased as well, falling from 3.86 on the MDNet model to 0.153. Overall, the findings described in suggest that data augmentation may help lessen the likelihood of overfitting and boost the effectiveness of models. Data augmentation makes the model more resilient by allowing it to learn several modified pictures from the original, smaller dataset. Applications in reality are quite likely to make use of such pictures, especially those involving one-of-a-kind human cells.

Table 6 shows how the accuracy of a DL model might vary depending on the learning rate used. Finding the best learning rate for a certain activity generally requires some trial and error in practice.

Table 5 The proposed model and the baseline model accuracy with and without augmentation

Model	*Without augmentation*		*With augmentation*	
	Accuracy	*Loss*	*Accuracy*	*Loss*
MDNet	98.21	1.5	98.9	0.153
CNN	99.01	2.0	99.86	0.231
VGG16	95.62	3.2	96.6	0.250
Resnet 50	94.32	3.5	95.12	0.182
AlexNet	95.12	3.8	96.84	0.265
EfficientNet	96.65	2.6	97.68	0.362

If the learning rate is excessively high, the model will make drastic changes to its settings after each iteration, which might lead to unpredictable results. Overshooting the optimal solution might result in poor convergence in the model. During training, the model's accuracy may fluctuate widely, making it difficult to zero in on the optimal answer. The

histopathological slide in its whole is a treasure trove of data. Pathologists spend a lot of time and effort doing manual analyses of histopathological pictures. Rapid and accurate assessment is now not possible due to the rising number of cancer patients. We propose a computer system for the accurate automated processing of histopathology pictures, which overcomes this current constraint. Our methodology generalises well beyond breast cancer.

Table 6 Accuracy of a DL model based on learning rate

Learning rate	*Accuracy*
0.0001	0.99
0.001	0.97
0.01	0.95
0.1	0.92
0.5	0.88
1.0	0.85

Modern neural networks are trained using learning rate decay (lrDecay). There is an initial burst of learning, followed by many periods of decline. Its usefulness for optimisation and generalisation has been practically shown. Decay rate (‘0.1’, ‘0.001’) clearly outperformed other decay rates on both sets of data. Both ‘0.01’ and ‘0.0001’ decay rates resulted in the lowest performance from the models.

6 Limitation and feature work

DL approach using MDNet-53 for RCC grading has exhibit promising results and it is also important to acknowledge potential limitations. The proposed approach may perform well on the specific dataset it was trained on but may struggle to generalise to unseen data or data from different sources. The diversity of RCC cases in real-world scenarios can be challenging to capture comprehensively.

The future work for an efficient DL approach using a MDNet-53 for automatic grading of RCC can involve areas of improvement and expansion. Here are potential directions for future research: data augmentation and diversity, multi-modal integration and explainable AI for more effective and efficient grading.

7 Conclusions

As far as we are aware, this is the first study to focus on radiological pictures (CT, MRI, etc.) for the purpose of grading kidney malignant tumours. Pathologists can benefit greatly from the suggested system, as it can help them implement an automated or semi-automatic cancer grading system, which can help alleviate some of the issues that arise from human grading. This study concludes by introducing a unique DL ensemble technique, MDNet, for the classification of RCC from kidney histopathology pictures. For optimal patient care, correct diagnosis and treatment planning of RCC, the most prevalent type of kidney cancer, are essential. Significant advancements have been made in the detection and classification of cancer in medical pictures with the use of DL

algorithms. In order to improve the accuracy and consistency of RCC classification, the proposed MDNet integrates multiple CNN models inside an ensemble architecture. In order to recognise RCC patterns on both the local and global scales, each CNN is tailored to extract certain features from the histopathological pictures. While a scale-matching technique helps choose appropriate scales for recognising small items, MDNet's fusion module gives extra knowledge about the targets. Extensive experimental studies were done to test the efficacy of the proposed method utilising a reference dataset of kidney histopathology pictures. To facilitate extensive testing and validation, the dataset included a wide range of RCC cases in addition to non-malignant kidney tissues. The outcomes prove that the ensemble DL approach provides superior performance in RCC classification compared to both individual CNN models and more conventional machine learning approaches. The MDNet method is highly successful, as evidenced by its 98.9% overall classification accuracy, 98.2% sensitivity, and 98.7% high classification specificity in differentiating tissues. The ensemble technique enhances stability and generalisability by combining data from different CNN models, and it shows promise as a diagnostic and therapeutic aid. Using DL techniques for RCC classification, the offered study advances the field of medical picture analysis. Improved patient outcomes may be possible because to more precise diagnosis and treatment planning made possible by the MDNet method. More research in this direction might increase the accuracy and timeliness of diagnostic tests, resulting to more effective treatment for RCC and better health outcomes for patients.

There are a number of potential paths forward for automatically grading kidney carcinoma with DL. Combining radiological imaging (CT, MRI) with genetic data can improve the recital of DL models. The accuracy and reliability of kidney cancer grading might be improved with the help of future research aimed at creating models that can efficiently include numerous types of data. In the future, it may be necessary to compile large, varied, and thorough databases of kidney cancer cases of differing stages. More robust and generalisable models may be trained using these datasets.

References

Alakwaa, W., Nassef, M. and Badr, A. (2017) 'Lung cancer detection and classification with 3D convolutional neural network (3DCNN)', *International Journal of Advanced Computer Science and Applications (IJACSA)*, Vol. 8, No. 8, pp.409–417.

Asteris, P.G., Gavriilaki, E., Touloumenidou, T., Koravou, E-E., Koutra, M., Papayanni, P.G. et al. (2022) 'Genetic prediction of ICU hospitalization and mortality in COVID-19 patients using artificial neural networks', *Journal of Cellular and Molecular Medicine*, Vol. 25, No. 5, pp.1445–1455.

Babu, J.C., Khan, M., Nuka, M.R. and Nagaraju, C.H. (2023) 'Detection and classification of COVID-19 using supervised deep learning on MRI images', *International Journal of Bioinformatics Research and Applications*, Vol. 19, No. 4, pp.233–251, DOI: 10.1504/IJBRA.2023.135362.

Bilaloglu, S., Wu, J., Fierro, E., Sanchez, R.D., Ocampo, P.S., Razavian, N., Coudray, N. and Tsirigos, A. (2019) *Efficient Pan-cancer Whole-slide Image Classification and Outlier Detection Using Convolutional Neural Networks*, bioRxiv.

Bouteldja, N., Klinkhammer, B.M., Bülow, R.D. et al. (2021) 'Deep learning based segmentation and quantification in experimental kidney histopathology', *J. Am. Soc. Nephrol.*, Vol. 32, No. 1, pp.52–68.

Bueno, G., Fernandez-Carrobles, M.M., Gonzalez-Lopez, L. and Deniz, O. (2020) 'Glomerulosclerosis identification in whole slide images using semantic segmentation', *Comput. Methods Programs Biomed.*, Vol. 184, p.105273.

Bukowy, J.D., Dayton, A., Cloutier, D., Manis, A.D., Staruschenko, A., Lombard, J.H., Woods, L.C.S., Beard, D.A. and Cowley Jr., A.W. (2018) 'Region-based convolutional neural nets for localization of glomeruli in trichrome-stained whole kidney sections', *J. Am. Soc. Nephrol.*, Vol. 29, No. 8, pp.2081–2088.

Bulten, W., Pinckaers, H., van Boven, H. et al. (2020) 'Automated deep learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study', *Lancet Oncol.*, Vol. 21, No. 2, pp.233–241.

Byeon, S.j., Park, J., Cho, Y.A. et al. (2022) 'Automated histological classification for digital pathology images of colonoscopy specimen via deep learning', *Sci. Rep.*, Vol. 12, p.12804 [online] https://doi.org/10.1038/s41598-022- 16885-x.

Chanchal, A.K., Lal, S., Kumar, R., Kwak, J.T. and Kini, J. (2023) 'A novel dataset and efficient deep learning framework for automated grading of renal cell carcinoma from kidney histopathology images', *Scientific Reports*, Vol. 13, No. 1, p.5728.

Chaudhari, S., Mithal, V., Polatkan, G. and Ramanath, R. (2021) 'An attentive survey of attention models', *ACM Trans. Intell. Syst. Technol. (TIST)*, Vol. 12, No. 5, pp.1–32.

Chen, C.L., Chen, C.C., Yu, W.H., Chen, S.H., Chang, Y.C., Hsu, T.I., Hsiao, M., Yeh, C.Y. and Chen, C.Y. (2020) 'An annotation-free whole-slide training approach to pathological classification of lung cancer types by deep neural network', *Nat. Commun.*, Vol. 12, p.1193.

Cireşan, D.C., Giusti, A., Gambardella, L.M. and Schmidhuber, J. (2012) 'Deep neural networks segment neuronal membranes in electron microscopy images', *Proceedings of the 25th International Conference on Neural Information Processing Systems – Volume 2, NIPS'12*, Curran Associates Inc., Red Hook, NY, USA, pp.2843–2851.

Echle, A., Grabsch, H.I., Quirke, P. et al. (2020) 'Clinical-grade detection of microsatellite instability in colorectal tumors by deep learning', *Gastroenterology*, Vol. 159, No. 4, pp.1406–1416.e11.

Eknoyan, G., Lameire, N., Eckardt, K., Kasiske, B., Wheeler, D., Levin, A., Stevens, P., Bilous, R., Lamb, E., Coresh, J. et al. (2013) 'KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease', *Kidney Int.*, Vol. 3, No. 1, pp.5–14.

Faridi, P., Danyali, H., Helfroush, M.S. and Jahromi, M.A. (2016) *Cancerous Nuclei Detection and Scoring in Breast Cancer Histopathological Images*, CoRR abs/1612.01237 [online] http://arxiv.org/abs/1612.01237.

Gallego, J., Pedraza, A., Lopez, S., Steiner, G., Gonzalez, L., Laurinavicius, A. and Bueno, G. (2018) 'Glomerulus classification and detection based on convolutional neural networks', *J. Imaging*, Vol. 4, No. 1, p.20.

Gao, Z., Puttapirat, P., Shi, J. and Li, C. (2020) 'Renal cell carcinoma detection and subtyping with minimal point-based annotation in whole-slide images', in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, Lima, Peru, 4–8 October, pp.439–448.

Gavriilaki, E., Asteris, P.G., Touloumenidou, T., Koravou, E-E., Koutra, M.m Papayanni, P.G., Karali, V. et al. (2021) 'Genetic justification of severe COVID-19 using a rigorous algorithm', *Clinical Immunology*, Vol. 226, DOI [online] https://doi.org/10.1016/j.clim.2021.108726.

Gu, Y., Ruan, R., Yan, Y., Zhao, J., Sheng, W., Liang, L. and Huang, B. (2022) 'Glomerulus semantic segmentation using ensemble of deep learning models', *Arab. J. Sci. Eng.*, Vol. 47, pp.1–12.

Hermsen, M., de Bel, T., den Boer, M. et al. (2019) 'Deep learning-based histopathologic assessment of kidney tissue', *J. Am. Soc. Nephrol.*, Vol. 30, No. 10, pp.1968–1979.

Jang, H.J. and Cho, K.O. (2019) 'Applications of deep learning for the analysis of medical data', *Arch. Pharm. Res.*, Vol. 42, No. 6, pp.492–504 [online] https://doi.org/ 10.1007/s12272-019-01162-9.

Jayapandian, C.P., Chen, Y., Janowczyk, A.R. et al. (2021) 'Development and evaluation of deep learning-based segmentation of histologic structures in the kidney cortex with multiple histologic stains', *Kidney Int.*, Vol. 99, No. 1, pp.86–101.

Jung, H., Lodhi, B. and Kang, J. (2019) 'An automatic nuclei segmentation method based on deep convolutional neural networks for histopathology images', *BMC Biomed. Eng.*, Vol. 1 [online] https://doi.org/10.1186/s42490-019-0026-8.

Kather, J.N., Heij, L.R., Grabsch, H.I. et al. (2020) 'Pan-cancer image-based detection of clinically actionable genetic alterations', *Nat. Can.*, Vol. 1, pp.789–799.

Kocak, B. et al. (2018) 'Textural differences between renal cell carcinoma subtypes: machine learning-based quantitative computed tomography texture analysis with independent external validation', *European Journal of Radiology*, Vol. 107, pp.149–157.

Kovesdy, C.P. (2022) 'Epidemiology of chronic kidney disease: an update 2022', *Kidney Int. Suppl.*, Vol. 12, No. 1, pp.7–11.

Kumar, G.S., Premalatha, K., Maheshwari, G.U. and Kanna, P.R. (2023) 'No more privacy concern: a privacy-chain based homomorphic encryption scheme and statistical method for privacy preservation of user's private and sensitive data', *Expert Systems with Applications*, Vol. 234, DOI [online] https://doi.org/10.1016/j.eswa.2023.121071.

Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M. and Mahmood, F. (2021) 'Data-efficient and weakly supervised computational pathology on whole-slide images', *Nat. Biomed. Eng.*, Vol. 5, pp.555–570.

Magneta, C.S., Sundar, C. and Thanabal, M.S. (2023) 'Lung lobe segmentation and feature extraction-based hierarchical attention network for COVID-19 prediction from chest X-ray images', *The Computer Journal*, Vol. 66, No. 2, pp.508–522, DOI [online] https://doi.org/ 10.1093/comjnl/bxac136.

Mahanty, C., Kumar, R., Asteris, P.G. and Gandomi, A.H. (2021) 'COVID-19 patient detection based on fusion of transfer learning and fuzzy ensemble models using CXR images', *Applied Sciences*, Vol. 11, No. 23, DOI [online] https://doi.org/10.3390/app112311423.

Mohapatra, S., Satpathy, S. and Paikaray, B.K. (2023) 'A machine learning approach to assist prediction of Alzheimer's disease with convolutional neural network', *International Journal of Bioinformatics Research and Applications*, Vol. 19, No. 2, pp.141–150, DOI: 10.1504/ IJBRA.2023.132632.

Odeh, I., Alkasassbeh, M. and Almseidin, M. (2022) 'Diabetic retinopathy detection method using artificial neural network', *International Journal of Bioinformatics Research and Applications*, Vol. 18, No. 4, pp.300–317, DOI: 10.1504/IJBRA.2022.126740.

Pan, X., Li, L., Yang, H., Liu, Z., Yang, J., Zhao, L. and Fan, Y. (2017) 'Accurate segmentation of nuclei in pathological images via sparse reconstruction and deep convolutional networks', *Neurocomputing*, Vol. 229, pp.88–99 [online] https://doi.org/10.1016/j.neucom.2016.08.103.; http://www.sciencedirect.com/science/article/pii/S0925231216313765.

Pantanowitz, L., Quiroga-Garza, G.M., Bien, L. et al. (2020) 'An artificial intelligence algorithm for prostate cancer diagnosis in whole slide images of core needle biopsies: a blinded clinical validation and deployment study', *Lancet Digit Health*, Vol. 2, No. 8, pp.e407–e416.

Polat, H. and Homay, D.M. (2019) 'Classification of pulmonary CT images by using hybrid 3D-deep convolutional neural network architecture', *Applied Sciences*, Vol. 9, No. 5, p.940.

Rajkomar, A., Oren, E., Chen, K. et al. (2018) 'Scalable and accurate deep learning with electronic health records', *NPJ Digit Med.*, Vol. 1, p.18 [online] https://doi.org/10.1038/s41746-018-0029-1.

Raman, S.P., Chen, Y., Schroeder, J.L., Huang, P. and Fishman, E.K. (2014) 'CT texture analysis of renal masses: pilot study using random forest classification for prediction of pathology', *Academic Radiology*, Vol. 21, No. 12, pp.1587–1596.

Shahin, M., Chen, F.F., Hosseinzadeh, A., Koodiani, H.K., Shahin, A. and Nafi, O.A. (2023) 'A smartphone-based application for an early skin disease prognosis: towards a lean healthcare system via computer-based vision', *Advanced Engineering Informatics*, Vol. 57, DOI [online] https://doi.org/10.1016/j.aei.2023.102036.

Stritt, M., Stalder, A.K. and Vezzali, E. (2020) 'Orbit image analysis: an open-source whole slide image analysis tool', *PLoS Comput. Biol.*, Vol. 16, No. 2, p.e1007313.

Temerinac-Ott, M., Forestier, G., Schmitz, J., Hermsen, M., Bräsen, J.H., Feuerhake, F. and Wemmert, C. (2017) 'Detection of glomeruli in renal pathology by mutual comparison of multiple staining modalities', in *Proceedings of the 10th International Symposium on Image and Signal Processing and Analysis*, Ljubljana, Slovenia, 18–20 September.

Tsuneki, M. and Kanavati, F. (2021) 'Deep learning models for poorly differentiated colorectal adenocarcinoma classification in whole slide images using transfer learning', *Diagnostics*, Vol. 11, p.2074 [online] https://doi.org/10.3390/diagnostics11112074.

Zhou, Y. and Yang, J. (2020) 'Chronic kidney disease: overview', in *Chronic Kidney Disease*, pp.3–12, Springer, Singapore.