Please cite the Published Version

Maqsood, Faiqa ![ID], Zhenfei, Wang, Ali, Muhammad Mumtaz, Qiu, Baozhi, Rehman, Naveed Ur, Sabah, Fahad, Mahmood, Tahir, Din, Irfanud and Sarwar, Raheem ![ID] (2024) Artificial Intelligence-Based Classification of CT Images Using a Hybrid SpinalZFNet. Interdisciplinary Sciences: Computational Life Sciences. ISSN 1913-2751

ORIGINAL RESEARCH ARTICLE

Artificial Intelligence-Based Classification of CT Images Using a Hybrid SpinalZFNet
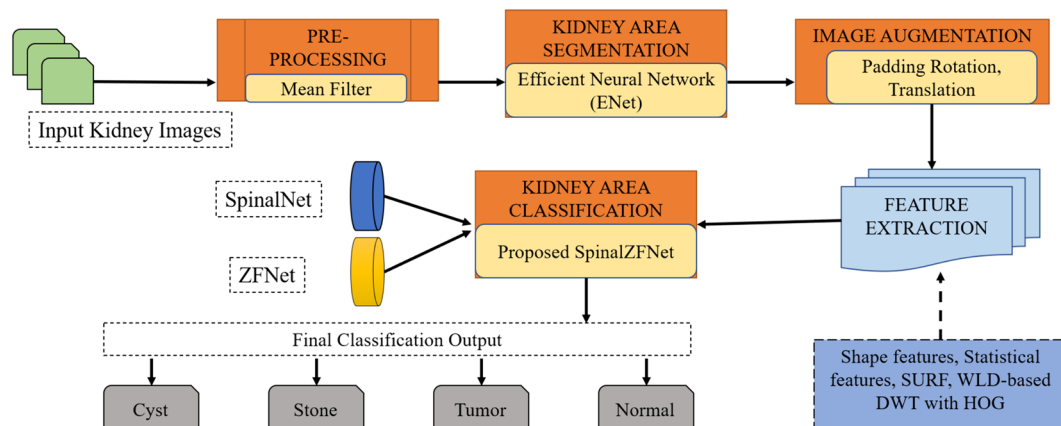
Faiqa Maqsood[1] · Wang Zhenfei[1] · Muhammad Mumtaz Ali[1] · Baozhi Qiu[1] · Naveed Ur Rehman[1] · Fahad Sabah[2] · Tahir Mahmood[3] · Irfanud Din[4] · Raheem Sarwar[5]

Abstract

The kidney is an abdominal organ in the human body that supports filtering excess water and waste from the blood. Kidney diseases generally occur due to changes in certain supplements, medical conditions, obesity, and diet, which causes kidney function and ultimately leads to complications such as chronic kidney disease, kidney failure, and other renal disorders. Combining patient metadata with computed tomography (CT) images is essential to accurately and timely diagnosing such complications. Deep Neural Networks (DNNs) have transformed medical fields by providing high accuracy in complex tasks. However, the high computational cost of these models is a significant challenge, particularly in real-time applications. This paper proposed SpinalZFNet, a hybrid deep learning approach that integrates the architectural strengths of Spinal Network (SpinalNet) with the feature extraction capabilities of Zeiler and Fergus Network (ZFNet) to classify kidney disease accurately using CT images. This unique combination enhanced feature analysis, significantly improving classification accuracy while reducing the computational overhead. At first, the acquired CT images are pre-processed using a median filter, and the pre-processed image is segmented using Efficient Neural Network (ENet). Later, the images are augmented, and different features are extracted from the augmented CT images. The extracted features finally classify the kidney disease into normal, tumor, cyst, and stone using the proposed SpinalZFNet model. The SpinalZFNet outperformed other models, with 99.9% sensitivity, 99.5% specificity, precision 99.6%, 99.8% accuracy, and 99.7% F1-Score in classifying kidney disease.

Graphical Abstract

Extended author information available on the last page of the article

1 Introduction

Worldwide, there has been an apparent increase in acute syndromes and infections, necessitating suitable clinical procedures to detect and treat these diseases immediately. If these diseases are left untreated, they will lead to serious health issues and death, resulting in a significant load on the healthcare system. The severity and mortality rate of acute diseases are much higher than that of infectious diseases [1]. The kidney is an abdominal organ that carries out the process of filtering the excess water and waste from the blood. Globally, over one-tenth of the population [2] is affected by the abnormalities that are developed in the kidneys, which also have the potential to cause an adverse effect on health. The precise cause of these abnormalities has not been completely recognized. However, several factors, such as lifestyle, environment, and genetics, have been determined to contribute to kidney abnormality in recent years [3]. These abnormalities are found to be commonly occurring in all individuals, irrespective of gender or age. Many abnormalities, including hydronephrosis, cysts, and stones, can affect the kidneys. While these conditions can be cured, if left untreated, they may lead to chronic kidney diseases such as cardiorenal syndrome, uremia, or kidney tumors [4]. Kidney cancer is one of the most common types of cancer worldwide, characterized by the abnormal growth of cells in the kidney. Kidney stones, which affect 12% of the population, are formed by the concretions of crystals. Cysts, on the other hand, are fluid-filled pockets surrounded by thin walls on the kidney surface [5].

As kidney abnormalities become more severe, symptoms such as swollen feet and legs, blood in urine, and pain in the back or sides may occur. Abnormalities, such as kidney tumor (renal cell carcinoma), kidney stones (nephrolithiasis), and cysts, can impair the normal functioning of the kidney [3]. The presence of tumor abnormalities can progress into severe conditions, such as chronic kidney diseases (CKD), a side effect of repeated stone abnormalities. The individuals with a high risk of stones are associated with an imminent risk of CKD and thus require dialysis. Hence, it is essential to identify kidney diseases at the primary stages to avoid serious complications [6]. Accordingly, the present diagnostic methods and patient management can be improved using Artificial Intelligence (AI) systems, which perform localization, classification, and quantification of kidney abnormalities based on images. Various imaging techniques, like positron emission tomography (PET), magnetic resonance imaging (MRI), computed tomography (CT), ultrasonography, X-rays, and other urography methods are used for capturing kidney images. Among these, CT scans are widely employed in detecting kidney diseases. However, every modality has strengths and weaknesses, resulting in distinct technical challenges while applying them [7]. Early detection is significant in avoiding severe complications at a later period. However, few radiologists and nephrologists are available worldwide, making timely detection challenging [5]. Further, accurate disease detection depends on the nephrologists' experience in the manual examination. This process is highly subjective, error-prone, and time-consuming, thus requiring the use of the computer-aided diagnosis (CAD) technique for detection [8, 9].

Identifying kidney diseases is a multidisciplinary domain that experts in information technology investigate to assist medical professionals by creating a model of the biological process occurring in kidneys and generating consistent diagnostic outcomes. In recent years, an increased occurrence of imaging datasets has been noted due to the ability to store massive data and the coincidence of sophisticated machine learning (ML) and deep learning (DL) methods [10]. Recently, ML techniques have been employed to analyze diseases efficiently by discovering the people at risk in the early stages. ML-based algorithms are currently the most extensively used approach to kidney disease detection [11]. These techniques learn and improve when vast amounts of data are integrated to learn the decision rules explicitly. DL approaches have also contributed to the sophisticated progress in AI, and these methods involve tuning multi-layer artificial neural networks (ANN) on massive databases [12, 13]. The incorporation of DL and computer vision schemes has been successfully applied in various medical domains, such as lesion detection [14] classification [15], and imaging segmentation. Medical professionals use robust approaches to detect cancer and cardiac diseases [16]. Recently, convolutional neural networks (CNNs), a DL approach, have enhanced the accuracy of various computer vision processes, like semantic segmentation, object detection, and image recognition. The CNNs can also capture complex tissue patterns successively and have been extensively applied in biomedical imaging for segmentation and classification processes [17].

The contribution of this research is as follows:

i) A hybrid SpinalZFNet technique is proposed to classify kidney disease accurately from CT images. The SpinalZFNet was devised as a combination of SpinalNet and Zeiler Fergus Network (ZFNet).

ii) Preprocess the images using a median filter to remove the noise from the data. The preprocessed image is segmented using an Efficient Neural Network (ENet).

iii) For feature extraction, different methods are used, such as weber local descriptor (WLD) based discrete wavelet transform (DWT) with histogram of oriented gradients (HOG) extracted used for feature extractors. The kidney

disease is finally classified from the extracted features into normal, cyst-tumor, and stone using the designed SpinalZFNet model.

iv) The performance of SpinalZFNet is evaluated based on the following: Sensitivity, Specificity, Precision, Accuracy, F1-Score, and Confusion matrix.

v) A comparative analysis was conducted comparing VGG + DN with KNN, DenseAUXNet201, MLP-ANN, and DRDC. The proposed model achieved a better 99.8% accuracy, 99.9% sensitivity, 99.5% specificity, 99.6% precision, and 99.7 F1-Score than the state-of-the-art models.

The paper workflow is as follows: Sect. 2 describes the related work overview. Section 3 presents detailed explanations of our proposed method. Section 4 describes the experimental datasets, evaluation metrics, and results. Section 5 discusses our findings. The conclusion is provided in Sect. 6.

2 Literature Review

This section summarizes the literature on different prevailing methods used for kidney cancer detection using images. Rajinikanth et al. [1] developed Visual Geometry Group (VGG) + DenseNet (DN) with K-nearest neighbor (KNN) to accurately classify renal CT images from healthy cancer classes. This model effectively removed the artefacts from the CT slices using a threshold filter-supported pre-processing method to obtain high performance. This approach produced high recognition accuracy with minimal computational complexity, but it required manual verification of the threshold while implementing the threshold filter. Mahmud et al. [7] designed a DenseNet201-based approach with auxiliary losses (DenseAUXNet201) for kidney cancer classification using CT images. This approach determined the cancer stage, the tumor volume, and standard demographic features to determine suitable surgical procedures. The DenseAUXNet201 attained a minimal count of missed cases, and further, the output was not biased by the kind of cancer. Moreover, the data augmentation did not completely address the problem of data imbalance, which resulted in no significant improvement in performance. Cai et al. [18] utilized an improved version of AlexNet that incorporated fused features. Their unique approach mitigated overfitting and demonstrated impressive performance, even with smaller datasets. This research paper presents a significant advancement in deep learning applications, particularly in limited data availability. While the model has shown increased resilience to overfitting, the study did not observe a substantial rise in classification accuracy. However, this should not discourage us. The researchers propose significantly boosting classification accuracy with additional modifications or complementary techniques. Shehata et al. [19] introduced multilayer perceptron-ANN (MLP-ANN) to determine whether the renal tumor is malignant or benign. It also decided on the sub-type of malignancy to provide optimal medical management. The surface complexity between the various tumors was effectively captured using morphological features, contributing to accurately determining malignancy status. However, the MLP-ANN did not consider including demographics, like sex and age, to improve performance.

Islam et al. [5] proposed a novel approach using a shifted window (Swin) Transformer with the visual geometry group (VGG)-16 architecture. This approach reduced training time while successfully capturing low-level features and enhancing their applications' outcomes. Integrating the Swin Transformer with VGG-16 represents a significant stride in efficiently processing complex visual data. Despite these advancements, the approach required high computational resources and involved many parameters. This high computational potential limitation, especially for deployment in environments with limited processing capabilities, points towards an area for future optimization. Patro et al. [16] introduced the deep kronecker neural network (DKN) model, exhibiting a noteworthy capacity to avoid overfitting, a common issue in deep learning models. The DKN was particularly effective in detecting even minuscule kidney stones, marking a substantial advancement in medical imaging. However, the efficiency of this approach was limited by the poor quality of the input images. The dependency on high-quality imaging indicates that inconsistency in imaging quality could affect the model's performance, emphasizing the need for consistent, high-quality input data for optimal results. Badawy et al. [8] developed the dualistic renal disease classification (DRDC) framework. This reliable system can produce promising outcomes even with limited data, making it useful in data-scarce situations. However, the DRDC framework did not explore the possibility of enhancing classifier performance by tuning classifier parameters with various optimization techniques. This gap suggests a potential area for future research where applying optimization techniques and hybrid models could improve the model's accuracy and efficiency.

Bhandari et al. [3] designed a lightweight CNN scheme to find abnormalities in the kidney, such as tumors, stones, and cysts. This method required minimal parameters and produced consistent results even with non-uniform sample distributions. However, this method was evaluated with limited CT images, which affected its performance. The issues endured by the available techniques for kidney cancer detection using images are as follows: VGG + DN with kNN proposed in [1], which required an artifact removal technique, and no image enhancement approach was considered to augment the accuracy of disease detection. The generalization of the dataset limited the DenseAUXNet201 [7], as the dataset

did not encompass all the information required to make the classifier robust. Further, the method should have considered developing a balanced dataset with numerous diverse cases for the approach to apply to a large population. In [19], DKN was a method that effectively detected kidney stones. However, the method suffered from high computational overhead, and no optimization technique was considered to reduce the computational effort. The DRDC framework developed in [8] achieved high classification accuracy with minimum misclassifications. Still, it was futile to consider combining various classifiers to improve detection efficiency and deploying the approach on the smartphone to make it more accessible. Several deep-learning techniques were proposed for classifying kidney diseases. However, these methods required high computational effort to produce superior performances. Moreover, insufficient data to train these networks affects the classification accuracy.

3 SpinalZFNet for Kidney Disease Classification Using CT Images

This research presents a hybrid DL model SpinalZFNet designed to classify kidney diseases from CT images. Initially, the input is acquired from the dataset and forwarded to the median filter. [20], which pre-processes the input images to denoise the input CT image. Later, kidney area segmentation is carried out with the help of the efficient neural network (ENet) [21] from the background image. Image augmentation is performed by padding, rotation, and translation to increase the image samples, thereby avoiding overfitting. This process is followed by feature extraction, where shape features, like area, solidity, eccentricity, perimeter, and primary axis length, as statistical features, such as mean (μ), entropy (H), correlation (C_r), and contrast C_t [22], speeded up robust features (SURF) [23], weber local descriptor (WLD) [24] based discrete wavelet transform (DWT) [25] with Histogram of Oriented Gradients (HOG) [26] extricated. Finally, kidney disease classification into (i) Normal, (ii) Cyst, (iii) Tumor, and (iv) Stone is accomplished based on hybrid SpinalZFNet, which is the combination of SpinalNet [27] and ZFNet [28], where layers are modified.

3.1 Dataset Description

The CT Kidney dataset [5] is collected from patients who have already been diagnosed with kidney stones, normal, cyst, or tumor findings in Dhaka, Bangladesh, by picture archiving and communication system (PACS). The contrast and non-contrast studies are performed for the urogram and abdomen to ensure a more in-depth capture of kidney details. The finalized dataset contains 12,446 distinct images. A breakdown of images includes 3709 images of

cysts, 5077 of normal kidneys, 1377 of kidney stones, and 2283 showcasing tumors.

3.2 Image Acquisition

The classification of kidney diseases is performed initially by acquiring CT images from the dataset.

$$K = \{K_1, K_2, K_3, K_4, \dots, K_l\} \tag{1}$$

Here are the CT images provided as input, which are donated as K_l.

3.3 Image Pre-Processing

The naturally available noises in the input CT image K_l are effectively removed by pre-processing using a median filter [20]. It is a nonlinear method that helps preserve useful information and remove noises in CT images by considering statistics. During pre-processing using a median filter, the actual pixel grey value is converted into grey values of pixels. The median is calculated by sorting all the pixel values of the neighborhood window into numerical order and replacing the pixel considered with the middle (median) pixel value. The median is calculated by the expression as follows:

$$P_l(u, v) = K_l(k, l) \tag{2}$$

Here, the input CT image given to the median filter is signified as $K_l(k, l)$, and $P_l(u, v)$ denotes the output CT image obtained while pre-processing using the median filter, which is further fed into ENet for image segmentation.

3.4 Kidney Area Segmentation Using Enet

The pre-processed CT image P_l is segmented from the kidney area from the background images using Enet [21]. The Enet is highly utilized for performing low latency operations, which also perform efficiently and require fewer parameters. In general, the view of Enet is constructed on a highly efficient convolutional network architecture tailored for rapid processing without compromising accuracy. Furthermore, Enet integrates bottleneck modules to minimize computational costs while maintaining robustness in segmentation tasks. Moreover, if the bottleneck is downsampling, the bottleneck modules and the max pooling to the main branch are considered. The activations are zero-padded to match the total feature maps. The model performs full, regular, or dilated convolutions with filters, but sometimes, the convolutional operations are replaced with asymmetric convolutions.

A single block is presented at the initial stage of ENet, where the bottleneck blocks are shown in stage 1. Moreover,

stages 2 and 3 have the same structure, but downsampling the input is not performed in stage 3. Stages 4 and 5 belong to the decoder, whereas the other initial three stages are encoders. The overall memory operations and kernel calls are reduced by not utilizing bias terms in any projections. Also, batch normalization is performed among each non-linear convolutional layer. In the decoding phase, max pooling is substituted with max unpooling, while padding gives way to spatial convolution devoid of bias. Pooling indices were not employed in the ultimate upsampling module due to the initial block processing of the input frame's three channels, in contrast to the final output featuring C feature maps representing object categories. Moreover, due to efficiency considerations, we opted to integrate only a singular full convolution at the network's end. This alone consumes a significant part of the decoder's processing duration. ENet's advanced features are essential for effectively segmenting the pre-processed CT images P_l.

3.5 Image Augmentation

The segmented CT image output S_o obtained using ENet allows image augmentation to avoid overfitting issues and increase diversity by randomly transforming the image samples. The different image augmentation approaches [5], like padding, rotation, and transformation, are performed in segmented CT image output S_o. The process carried out during augmentation is briefly enumerated further in detail. First, the padding adds pixels to the sides of the segmented CT image S_o. It is the total number of pixels added to the segmented image while processing among kernels of CNN, where the resultant padding augmented image is represented as W_o. Secondly, the rotation augments the segmented CT image S_o by altering its orientation. This involves pivoting the image around its center to analyze the direction of the image along the angles between $1°$ and $359°$. The term W_l represents the augmented CT image output obtained during rotation. Translation is applied for the movement of objects available in the segmented CT image S_o from one position to another. The images are preserved or randomized by translating them into black or white along X and Y direction or towards X or Y direction at the same time, where, during translation, the position bias is adjusted towards the up, down, right, and left directions. The resultant translation augmented CT image is represented as W. Therefore, $W = [W_1, W_2, W_3, \ldots, W_n]$ is the final augmented CT image obtained during the augmentation of the segmented CT image S_o.

3.6 Features Extraction

Advanced methods are used to extract pertinent information from the images during feature extraction.

Specifically, features such as mean (μ), entropy (H), correlation (C_r), and contrast C_t [22], SURF [23], WLD [29] based DWT [25] with HOG [26] are used for extraction of features from augmented CT images W. The extraction process performed is explained below.

3.6.1 SURF

The augmented CT image W is fed into SURF to extract textural features. The SURF is a local feature commonly used to locate interesting points and determine descriptors.

Localization of interest point: in general, the SURF detector is based on the Hessian matrix:

$$H(Z, \sigma) = \begin{bmatrix} L_{xx}(Z, \sigma) & L_{xy}(Z, \sigma) \\ L_{xy}(Z, \sigma) & L_{yy}(Z, \sigma) \end{bmatrix} \tag{3}$$

where, $L_{xx}(Z, \sigma)$ is the convolution of Gaussian second order of derivative in point Z, which is also similar for $L_{xy}(Z, \sigma)$ and $L_{yy}(Z, \sigma)$. The Hessian is determined for selecting interest point scale and location, whereas non-maximum suppression is applied for localizing the interest points in image and scale space. Moreover, the interpolation of image and scale space is performed finally to identify the local maxima of the appropriate Hessian matrix.

Descriptor interest point: in this phase, a unique orientation is assigned to the former, and image rotation is performed by gaining invariance by constructing a circular region around the detected point of interest. The orientation is computed using Haar wavelet responses in both x and y. The Haar wavelets are quickly computed via integral images, similar to the Gaussian second-order approximated box filters. The dominant orientation is estimated and included in the interest point information. In the next step, square regions are extracted around the point of interest for constructing the SURF descriptor. The spatial data is retained by splitting the windows into sub-regions, where Haar wavelets are extracted in each sub-region. The polarity information of image intensity changes is obtained by summing the absolute values. Therefore, the vector V used to underline the intensity patterns of the sub-regions is given in the following expression:

$$V = \left[\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y| \right] \tag{4}$$

where, d_x and d_y signifies the wavelet response over horizontal and vertical directions. The Haar wavelets are invariant to contrast and are also invariant to illumination bias while normalizing the resultant descriptor vector, and the extracted SURF feature is signified as J_1.

3.6.2 WLD-Based DWT with HOG

The discrete wavelet transform (DWT) refers to linear transformation, a function of the data vector, the length of which is associated energy. In the wavelet transformation, two steps are used to isolate the function of different natural image subbands and isolate image wavelet coefficients. The expression used to compute the mean value of DWT coefficients is given in Eq. (5):

$$\mu_t = \text{mean(DWT)} \tag{5}$$

where, the determined inference value from the initial value is represented as μ_t. The DWT splits the augmented CT image W into three subbands, as Low-Low (LL) ϕ_1, Low-High (LH) ϕ_2, and High-High (HH) ϕ_3. The split subbands are separately fed into WLD for further augmented image extraction. Further, the WLD used a simple yet robust local texture descriptor for digital images, which Weber's Law inspired. It is composed of two components: differential excitation and orientation. The differential excitation of the present pixel is determined using the expression given below:

$$g(f_h) = \arctan\left[\sum_{i=0}^{t-1} \frac{f_i - f_h}{f_h}\right] \tag{6}$$

Here, f_h signifies the present pixel. f_i represents the i-th neighboring pixel of f_h and the total amount of neighbor are represented as t. Moreover, the present pixels' gradient orientation component is determined by Eq. (7):

$$\phi(f_h) = \arctan\left[\frac{\phi^{(11)}}{\phi^{(10)}}\right] \tag{7}$$

where the output obtained from the two filters is assumed as $\phi^{(11)}$ and $\phi^{(10)}$. The subbands are obtained from DWT by filtering the image using the kernels depicted as follows:

$$Y_o = (-1, 0, 1) \text{ and } Y_s = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \tag{8}$$

Also, the orientation $w(p, q)$ as well magnitude $v(p, q)$ is determined using the expression using the Eq. (9) as follows:

$$w(p, q) = \sqrt{(A_x^2 + A_y^2)} \text{ and } v(p, q) = A_y/A_x \tag{9}$$

Here, the vertical and horizontal gradients are represented as A_y and A_x, and the final extracted feature obtained by WLD-based DWT with HOG feature extractor is signified as J_2. Moreover, the extracted textural features are given by $J = [J_1, J_2]$, which further allows the extraction of the most significant vector features using shape and statistical feature extractors.

3.6.3 Shape Features

Shape features, like area, perimeter, major axis length, minor axis length, and convex hull, are used to calculate the shape and find its value. Details of shape features are provided as follows: The shape is extracted using compactness, which is calculated by using the ratio of perimeter, and the extracted area feature is signified as H_1. The extent to which the shape is concave or convex is determined using solidity [30], which is computed by:
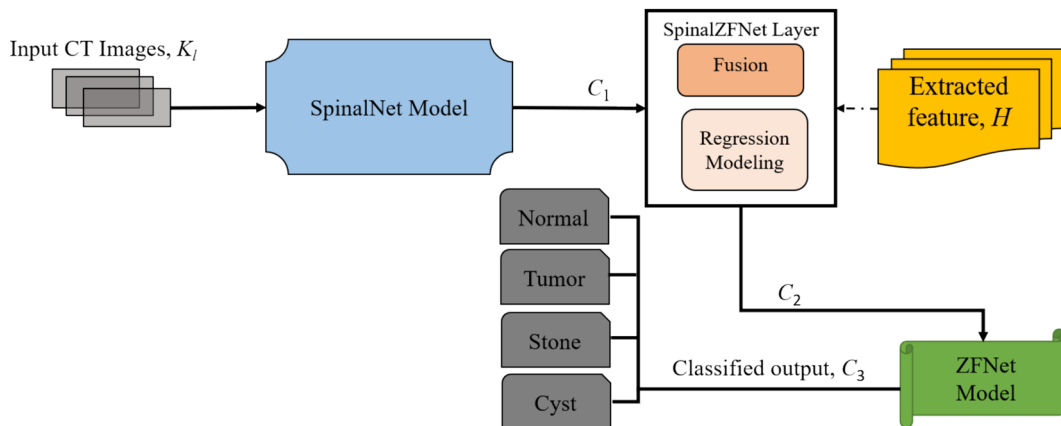
$$H_2 = \frac{\alpha}{\beta} \tag{10}$$

Fig. 1 Block diagram of SpinalZFNet used for kidney disease classification

where the object area is represented as α, the convex hull is an outside line that covers all the points and is signified as β, and the extracted solidity feature is signified as H_2. Eccentricity is the calculation of the length of the central axis as a proposition to the length of the minor axis [30], given by expression as:

$$H_3 = \frac{\zeta}{\delta} \tag{11}$$

Here, the extracted eccentricity feature is symbolized as H_3, the minor axis is signified as ζ and the major axis is denoted as δ. Perimeter is the summation of pixels at all the sides of an image's shape. This distance, represented by Eq. (12), measures the shape's size in the respective dimension:

$$H_4 = \sqrt{(a_2 - a_1)^2 + (b_2 - b_1)^2} \tag{12}$$

where, the endpoints are represented as (a_1, b_1) and (a_2, b_2) and the extracted major axis length is signified as H_4. Thus, the extracted shape features are finally given by the expression: $H_{shape} = \{H_1, H_2, H_3, H_4\}$.

3.6.4 Statistical Features

The different statistical feature extractors, like mean, entropy, correlation, and contrast are extracted from the augmented CT images. The extraction process performed is briefly described as follows: The pixels available in the image under a certain dimension are termed mean [5], which is computed using:

$$H_5 = \frac{1}{R} \sum_{c=i}^{R} I_c \tag{13}$$

Here, the image pixel is signified as I, and H_5 indicates the extracted mean feature. The pixel with high information is determined using entropy [5]. The formula is expressed as follows:

$$H_6 = \sum_{c,d} p(c,d) \log p(c,d) \tag{14}$$

$p(c,d)$ is the probability of occurrence of pixel value c, d in the image and H_6 signifies the extracted entropy feature. The correlation [22] is used to describe the correlation of a pixel with another pixel of the image, where the expression gives the correlation feature:
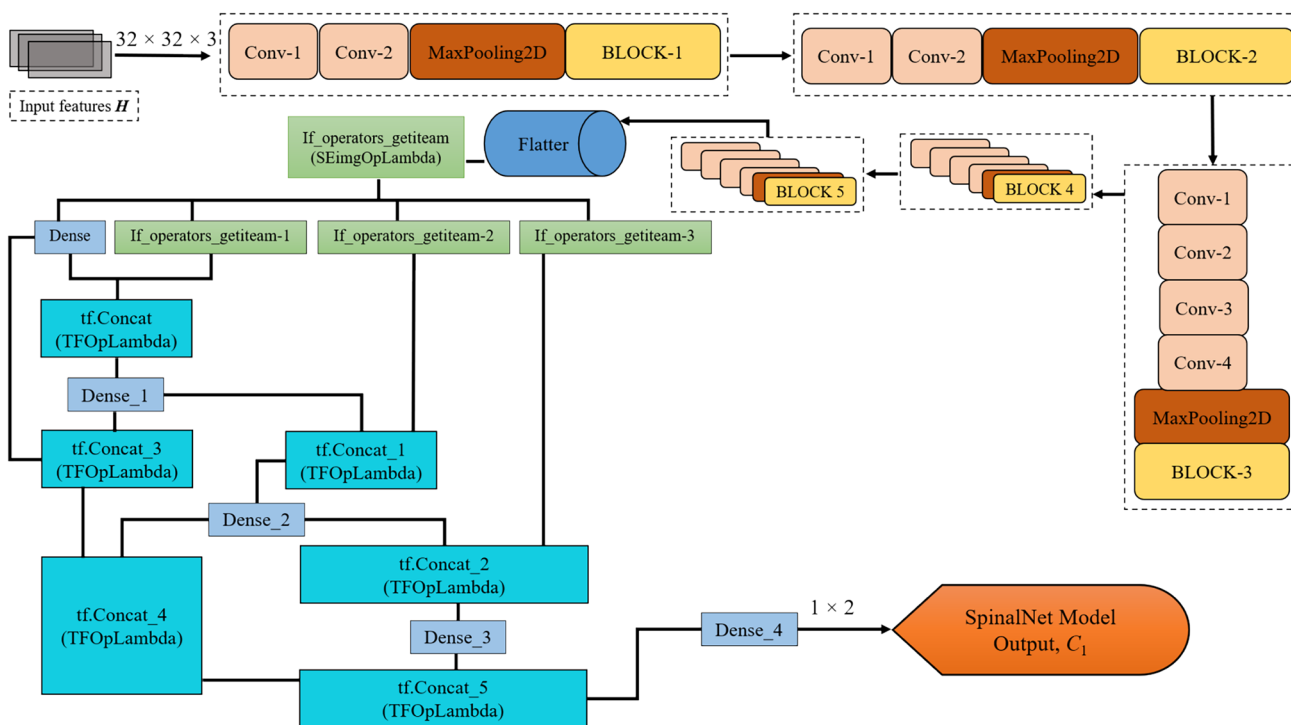

Fig. 2 Structure of SpinalNet Model

$$H_7 = \sum_{c,d} \frac{(c - \Psi_c)(d - \Psi_d)\eta_{c,d}}{\mathcal{L}_c \mathcal{L}_d} \tag{15}$$

where, c and d indicates section, Ψ signifies mean, the standard deviation is represented as $\mathcal{L}$, the Grey-Level Co-occurrence Matrix (GLCM) is symbolized as η, and the extracted correlation feature is given by H_7. Contrast is the extent reduced level arrangement of a neighborhood in the image and is given in Eq. (16) as

$$H_8 = \sum_{c,d} |c - d|^2 \eta_{c,d} \tag{16}$$

where the extracted contrast feature is signified as H_8. Here $H_{stat} = \{H_5, H_6, H_7, H_8\}$ is the final extracted shape given in the expression. Finally, all features are extracted further to classify the kidney disease.

3.7 Kidney Disease Classification Using SpinalZFNet

The abnormalities available in the kidney can potentially cause an adverse effect on health. Hence, it is necessary to detect the exact kidney disease to provide accurate treatment for kidney disease. Thus, the SpinalZFNet model is designed in this research to precisely classify kidney disease into normal, cyst, tumor, and stone to provide early diagnosis. Here, the SpinalZFNet model is initially processed by the SpinalNet model, which is trained to recognize complex patterns associated with kidney conditions. This preliminary stage acts as a filter, refining the input and preparing it for more detailed analysis. Then, the extracted feature H along with the output obtained from SpinalNet C_1 is sent to SpinalZFNet layer to obtain C_2. This layer facilitates the fusion of technologies from SpinalNet [27] and ZFNet [28], employing a fully connected (FC) [31] setup for regression modelling, which aims to identify relationships among the inputs and perform the fusion. This FC setup is particularly well-suited for tasks involving various integral and derivative calculations due to its dense network of node connections. It effectively integrates and extracts high-level features from the data, enhancing the model's capability to handle complex information processing. Consequently, the output C_2 obtained from the SpinalZFNet layer is sent to the ZFNet model for the final kidney disease classification. The final classified output of C_3 is obtained, which is classified as (i) Normal, (ii) Cyst, (iii) Tumor, and (iv) Stone by forwarding the output C_2 obtained from SpinalZFNet layer to ZFNet model. The ZFNet model component improved the model's flexibility, maintainability, and potential for fine-tuning. The ZFNet model is a pre-trained feature extraction module easily integrated into the SpinalNet model, using transfer learning principles to enhance the model's performance. The block diagram of SpinalZFNet used for kidney disease classification is displayed in Fig. 1.

3.7.1 SpinalNet Model

The SpinalNet model [27] is designed by considering the function of the human spinal cord, where the input K_l CT image are given to the SpinalNet continuously in a step-by-step format. The global and local output sends a modulated input version pass through each SpinalNet layer. The model structure contains input sub-layer, intermediate sub-layer, and output layer. The input is split and forwarded to multiple hidden layers known as intermediate network layers. The intermediate split layer of SpinalNet generally acquires two components, one from the previous intermediary layer split outcome and one from the present input split outcome. The weighted outputs of intermediate splits are passed and added with each layer's output split. The hidden layers of intermediate splits comprise two neurons, where the expression gives the output of the neuron in the hidden layer,

$$C_1 = \sum_{i=1}^{N} (w_i)q_i + b_i \tag{17}$$

Here, the threshold value of b_i represents the bias associated with the neuron in the hidden layer. This bias helps adjust the output and the weighted sum of the inputs to the neuron. The weights w_i correspond to the significance of each input q_i and the summation runs over all N inputs to that neuron. The depth and architecture of the SpinalNet model enable it to effectively identify intricate patterns and relationships within the input data, making it particularly suited for identification of kidney diseases once the input CT image is fed into the SpinalNet model. The structure of SpinalNet model is depicted in Fig. 2. The model integrates the outputs from various stages of the network, including the convolutional blocks (Conv-1, Conv-2, MaxPooling2D, BLOCK-1 to BLOCK-5) and the dense operators (Dense_1 to Dense_4). The convolutional blocks extract local features from the input data, while the dense operators process the flattened output from the convolutional blocks. The concatenation operations (tf.Concat_1 to tf.Concat_5) combine these feature representations at different levels, allowing the SpinalNet model to leverage multi-scale information.

3.7.2 SpinalZFNet Layer

The resultant output C_1 obtained from the SpinalNet model and the extracted features $\boldsymbol{H}$ is fed into SpinalZFNet layer for the fusion of SpinalNet [27] and

ZFNet [28]. The FC [31] is used for regression modeling in SpinalZFNet layer to perform fusion by determining the relationship between the given inputs. FC generally provides solutions to various integral and derivative tasks by establishing dense node connections. These layers integrate information, capturing high-level features from the input data. In the context of the provided equation, the fusion computes the weighted sum of features from different intervals or sources. Layers at p-th interval leverage this weighted approach to better capture the intricacies and patterns across multiple data points. Specifically, p is given in Eq. (18),

$$p = \sum_{I=1}^{2} \sum_{F=1}^{s} \sum_{G=1}^{t} H_{IFG} \times X_{IFG} \tag{18}$$

Here, the available total amount of features is indicated as t, s indicates total available data, the weight coefficient is signified as X. At $(p-1)$-th interval the expression gives the SpinalZFNet layer output,

$$p_1 = \sum_{E=1}^{4} H_{ST} \times X_E \tag{19}$$

where, the statistical feature is represented as H_{ST}. Moreover, the output of SpinalZFNet layer at $(p-2)$-th interval is designated as,

$$p_2 = \sum_{E=1}^{4} H_{SN} \times X_E \tag{20}$$

where H_{SN} signifies shape.

$$U = f(p, p_1, p_2) \tag{21}$$

Equation (21) is given by

$$C_2 = U \times p + \frac{1}{2} U \times p_1 + \frac{1}{6}(1-U) \times p_2 + \frac{1}{24}(1-U)(2-U) \times C_1 \tag{22}$$

Thus, applying the value of p, p_1 and p_2, Eq. (22) becomes.

$$C_2 = U \times \left[\sum_{I=1}^{2} \sum_{F=1}^{s} \sum_{G=1}^{t} H_{IFG} \times X_{IFG} \right] + \frac{1}{2} U \times \left[\sum_{E=1}^{4} H_{ST} \times X_E \right] + \frac{1}{6}(1-U) \left[\sum_{E=1}^{4} H_{SN} \times X_E \right] + \frac{1}{24}(1-U)(2-U) \times C_1 \tag{23}$$

Equation (23) is the SpinalZFNet layer's final resultant output, which is forwarded to the ZFNet model for kidney disease classification.
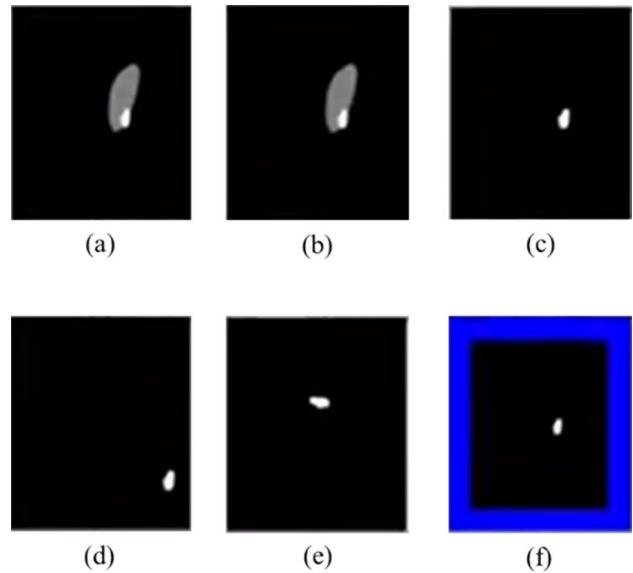

Fig. 4 Experimental image outcomes. **a** Input, **b** pre-processed, **c** segmented, **d** translation, **e** rotation, and **f** padding augmented CT image
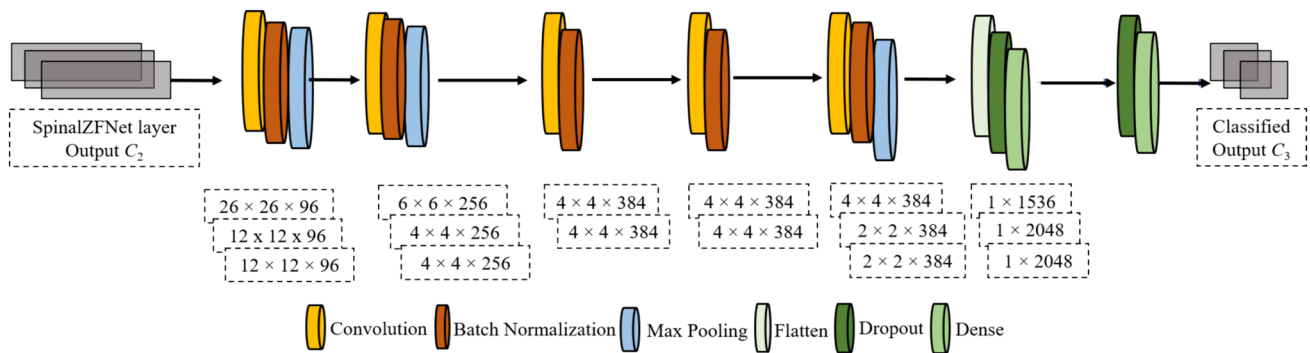

Fig. 3 Structure of ZFNet model

3.7.3 ZFNet Model

A type of CNN, the ZFNet model [28, 32] is introduced by visualizing the intermediate feature maps. The ZFNet structure consists of convolutional and fully connected (FC) layers, where the output is fed into the subsequent layer for further processing and refinement. This hierarchical structure ensures a progressive refinement of features as the data traverses through the network. The kernel map of the second layer is connected to the kernel of the third convolutional layer, and the fully connected layer neurons are connected to the neurons of the preceding layer. In addition, next to the first and second convolutional layers, the response-normalization layer is connected. Followed by the response-normalization layer and the fifth convolution, the max pooling layer is connected. Rectified Linear Units (ReLU) non-linearity is applied to the convolutional and fully connected layer output. The kidney disease classification process carried out in the ZFNet model is explained below.

The output obtained from the SpinalZFNet layer, C_2, is fed initially into a convolutional layer that forwards the output in a successive or sequential form via interconnected convolutional layers. Here, the input is convolved with kernels for the creation of feature maps, and the convolution operation performed in a convolutional layer is designated as:

$$\varpi = C_2 \vartheta \tag{24}$$

From the expression, the

$$L = t(\varpi) = t(C_2 \vartheta) \tag{25}$$

The activation function is signified as t, and the matrix is represented as L. The down-sampling of feature maps available in the channels is performed in the max-pooling layer. Moreover, the summation is performed for the determination of main features, and finally, a stride $J \times J$ is used to generate a reduced subset after the output obtained from the convolutional layer is fed into the max pooling layer, which is expressed as

$$o : L \rightarrow \theta$$

$$\mathrm{R}^{f^2 \times f_a} \rightarrow \mathrm{R}^{\frac{f^2}{J^2} \times f_a} \tag{26}$$

Here, the reduced subset is signified as θ and f_a indicates the pooling operation carried out in channels. The max-pooling operation is performed by considering the highest pooling value and is designated as

$$\theta = B_{\max}\big(t(C_2 \vartheta)\big) \tag{27}$$

Here, the max-pooling function is represented as $B_{\max}$ and is formulated as

$$B_{\max} = \Gamma_{\max}(O)O \tag{28}$$

where the identity matrix is signified as $\Gamma_{\max}(O) \epsilon \mathrm{R}^{\frac{f^2}{J^2} \times f^2}$

In the ZFNet model, the flattening layer is used for the conversion of feature maps into a set of elements before forwarding it to the fully connected layers

$$z = \big[v_1, v_2, \dots, v_f\big]$$

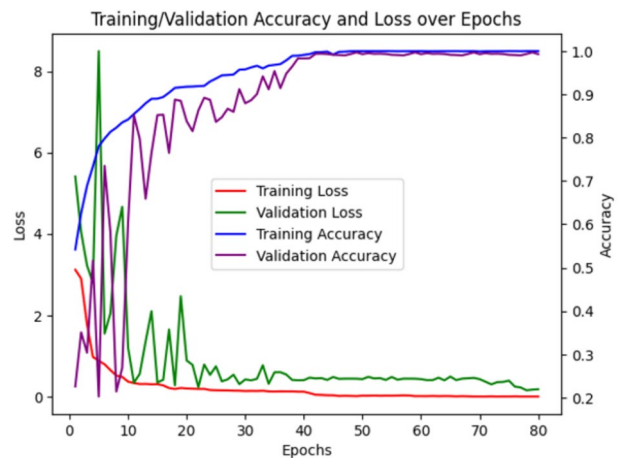$$= [\Gamma_1 v_1, \Gamma_2 v_2, \dots, \Gamma_f v_f] \tag{29}$$

Fig. 5 Training and validation accuracy and loss curves on the CT images of the SpinalZFNet model

Table 1 Details of the hyperparameters used in the SpinalZFNet

Hyperparameter	Variables	Randomly used
Optimizer	Adam, Nadam, SGD, RMSProp	Adam, SGD, RMSProp
Batch size	12, 16, 32, 64	32
Epochs	[55, 65, 100, 120, 150]	100
Learning rate	0.01, 0.001, 0.0001	Learning rate adjusted based on the training progress

where, the f-th max pooling layer output of the pooling layer and the flattening operation executed in the flattened layer is given in Eq. (30).

$$z = F[z] = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_f \end{bmatrix} = \begin{bmatrix} \Gamma_1 v_1 \\ \Gamma_2 v_2 \\ \vdots \\ \Gamma_f v_f \end{bmatrix} \qquad (30)$$

The dense layer performs linear operation by determining the relationship between the inputs. Thus, the resulting output is designated as

$$z = F[z] = F\Gamma_n N_n r\left(C_2 \vartheta_n\right) \qquad (31)$$

The expression gives the final output of CNN,

$$C_3 = G_k\left(D_k \dots G_2(D_2 G_1\left((D_1 G)\right) \dots \right) \qquad (32)$$

The weight and activation functions are denoted as D_k and G_k at the k-th FC layer, respectively. Activation function output is donated as

$$E_3 C_3 = E\left(D_k \dots G_1(D_1 \times \Gamma_2 N_2 [\Gamma_1 N_1 (C_2^{f_1} \vartheta_1)]^{f_2} \vartheta_2) \dots \vartheta_n\right) \qquad (33)$$

The activation function is indicated $E_3 C_3$ The structure of the output of the ZFNet model is presented in Fig. 3.

4 Result

This section discusses the results obtained by applying the SpinalZFNet model to classify kidney diseases. Our goal was to leverage the advantages of the spinal architecture and the ZFNet model to achieve a more precise and comprehensive classification. We conducted experiments that demonstrated its effectiveness. The SpinalZFNet was designed to classify kidney disease using the CT kidney dataset [5].

4.1 Experimental Image Results

The resultant images recorded by SpinalZFNet during kidney disease classification display in Fig. 4. The input CT image show in Fig. 4a, b depicts the pre-processed CT image, and Fig. 4c gives the segmented CT image. Moreover, Fig. 4d–f portray the translation, rotation, and padding augmented CT image. Given the dataset, such as kidney disease images, calculate the number of features each method extracts from a single image as follows:

(1) Speeded-up robust features (SURF): the image is of a kidney disease segment image based on a black background. SURF detects only a few key points due to the high contrast and small area of interest. SURF detects five key points.

With a descriptor length of 64, that would be 320 features (5 key points 64×64 descriptor length). (2) WLD-based DWT with HOG (Wavelet-Local Binary Patterns with Histogram of Oriented Gradients): after applying DWT, approximation coefficients (cA) and detail images (cH, cV, cD) are all considered for HOG feature extraction. With an HOG descriptor window size of 64×64 pixels, 8×8 cells, and 9 histogram bins per cell, get a feature vector of length $7 \times 7 \times 9 = 441$ from a single channel (due to the downscaling effect of DWT). (3) Shape features: given that the image's shape is simple, calculate one value each for area, perimeter, major axis length, minor axis length, and convex hull. This results in 5 shape features. (4) Statistical features: the image's region has some statistical measures: mean intensity, entropy (calculated from the histogram of pixel intensities), correlation, and contrast (from a gray-level co-occurrence matrix, (GLCM)). This extracts 4 statistical features.

4.2 Training and Implementation

This study trained the proposed SpinalZFNet model using 100 epochs on the Google Colab platform. The training was conducted on a single 12 GB NVIDIA Tesla K80 GPU with a maximum continuous usage limit of 12 h. The GPU mode was utilized for faster execution. However, the completion time of the training process depended on factors such as network speed and dataset size.

Table 1 presents a comprehensive overview of the hyperparameters used in the SpinalZFNet model. We employed random hyperparameters to train the proposed model. We utilized multiple optimizers such as Adam, SGD, and RMSProp for the random hyperparameters. We set the default learning rate to 0.001. During the training, we incorporated critical features such as a reduced learning rate, model checkpoint, and early stopping. The reduced learning rate helped adjust the learning rate dynamically based on the training progress. The monitoring of the validation accuracy was conducted, and if it did not exhibit improvement for five consecutive epochs, with a minimum change of 0.0001, the present learning rate was halved. This process continued until the last epoch.

We also utilized the model checkpoint callback to store the weights of the top-performing model, which was determined by its validation accuracy. Early stopping was employed to ascertain the optimal number of epochs required for the training process. After 50 epochs, if no performance improvement was observed, the training process was stopped to avoid overfitting. The batch size, another essential hyperparameter, was set to 32 in the SpinalZFNet model, as it yielded satisfactory results. Larger batch sizes allow for better parallelization, which may negatively affect the model's generalization ability. The study employed the SoftMax cross-entropy loss function, which is well-suited
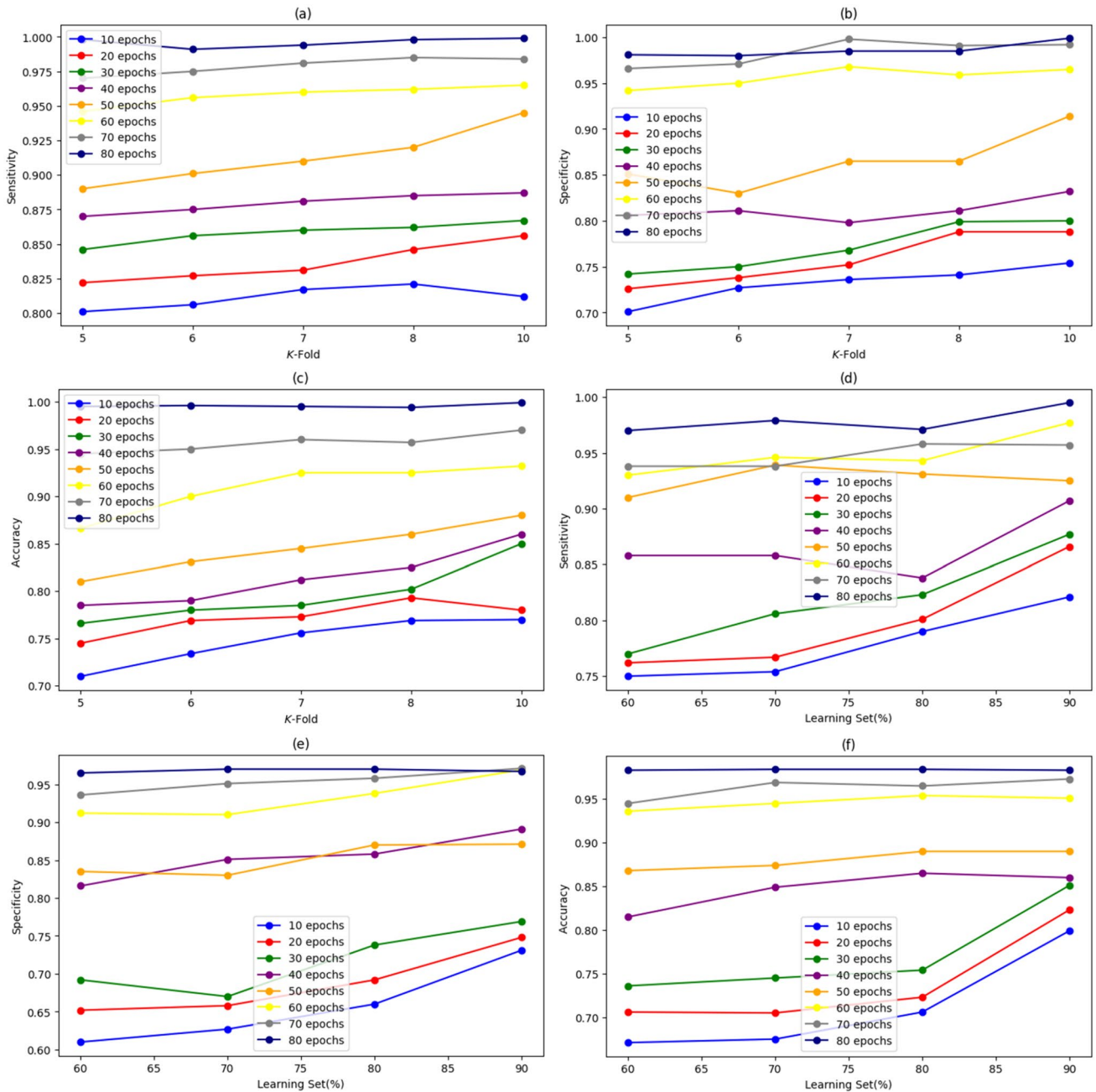
Fig. 6 Performance analysis of SpinalZFNet using *K*-fold (**a** sensitivity, **b** specificity, **c** accuracy) and learning set, (**d** sensitivity, **e** specificity, and **f** accuracy)

for addressing multiclass classification problems—quantifying the discrepancy between the output of the network and its corresponding labels.

Figure 5 presents the learning curves of the SpinalZFNet model on the CT dataset. The curves for training and validation, as well as the loss curves, show an effective convergence across the dataset. The training process was stopped after 80 epochs to avoid overfitting, as no improvement in performance was observed for five consecutive epochs. The

optimal number of epochs was utilized for training. This means that the data used for training the model is a good representation of the data used for validation. The fact that the training and validation curves are close to each other suggests that the model is reliable when tested. This provides evidence that the SpinalZFNet model is not overfitting.
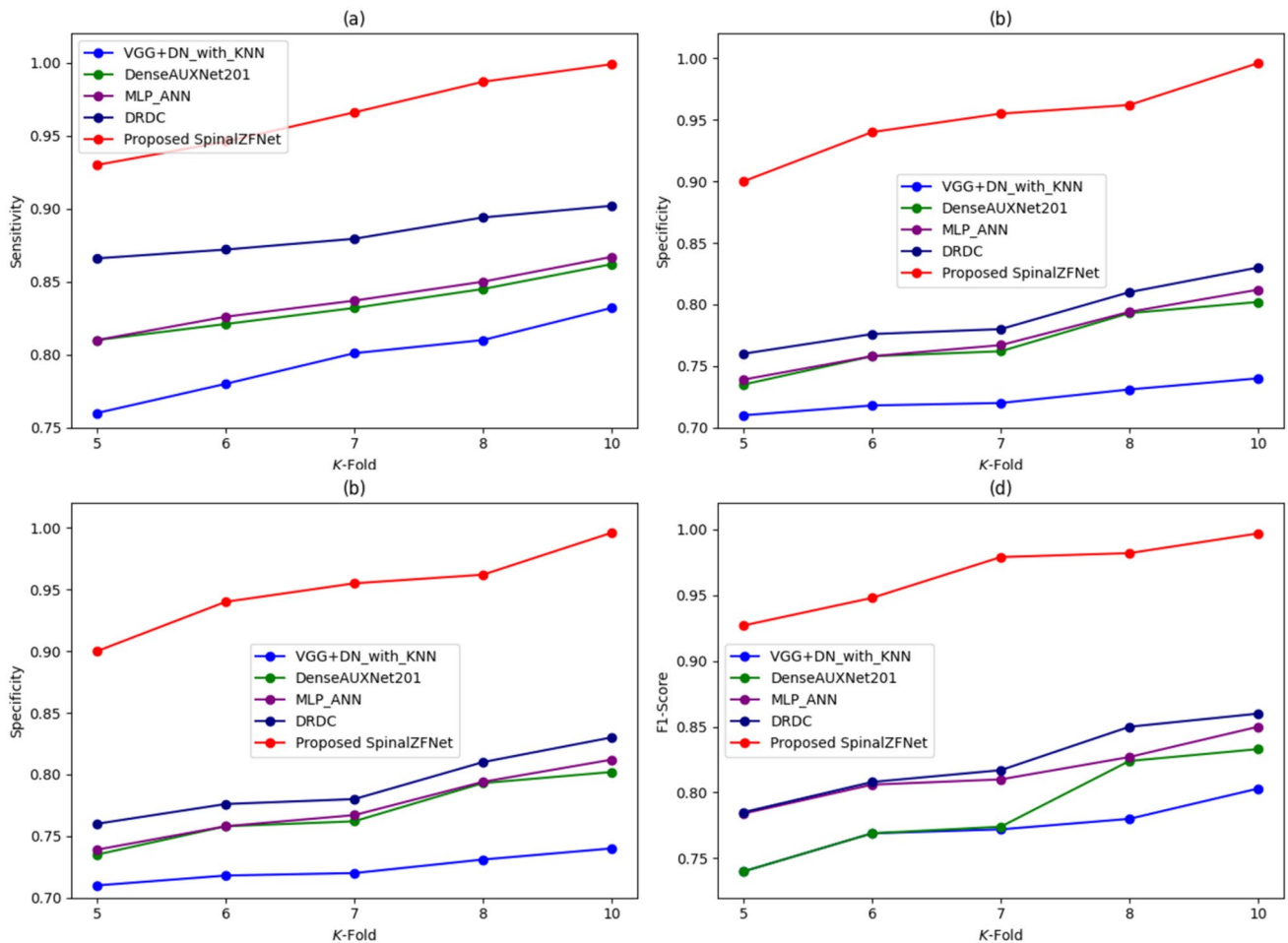
Fig. 7 Comparative analysis of SpinalZFNet using *K*-fold: sensitivity, specificity, accuracy, and F1-score

4.3 Performance Metrics

Various performance metrics [33, 34] are utilized to identify SpinalZFNet's superiority, including the classification performance, which are explained below.

Accuracy: is the relationship between the original and expected outcome obtained by SpinalZFNet during kidney disease conditions.

$$Accuracy = \frac{n_{TP} + n_{TN}}{n_{FP} + n_{FN} + n_{TP} + n_{TN}} \quad (34)$$

where $n_{TP}, n_{TN}, n_{FP}, n_{FN}$ represent the number of true positives, true negatives, false positives, and false negatives, respectively.

Sensitivity: the positive samples classified precisely by SpinalZFNet from the input total positive samples is called sensitivity, where the sensitivity is given by

$$Sensitivity = \frac{n_{TP}}{n_{TP} + n_{FN}} \quad (35)$$

Precision: the precision is defined as the positive samples falsely classified SpinalZFNet from the total positive samples and is defined as

$$Precision = \frac{n_{TP}}{n_{TP} + n_{FP}} \quad (36)$$

Specificity: the specificity is defined as the negative samples precisely classified by SpinalZFNet from the total input negative samples and is expressed by

$$Specificity = \frac{n_{TN}}{n_{TN} + n_{FP}} \quad (37)$$

F1-Score: Precision and Sensitivity are used by SpinalZFNet for F1-Score calculation.
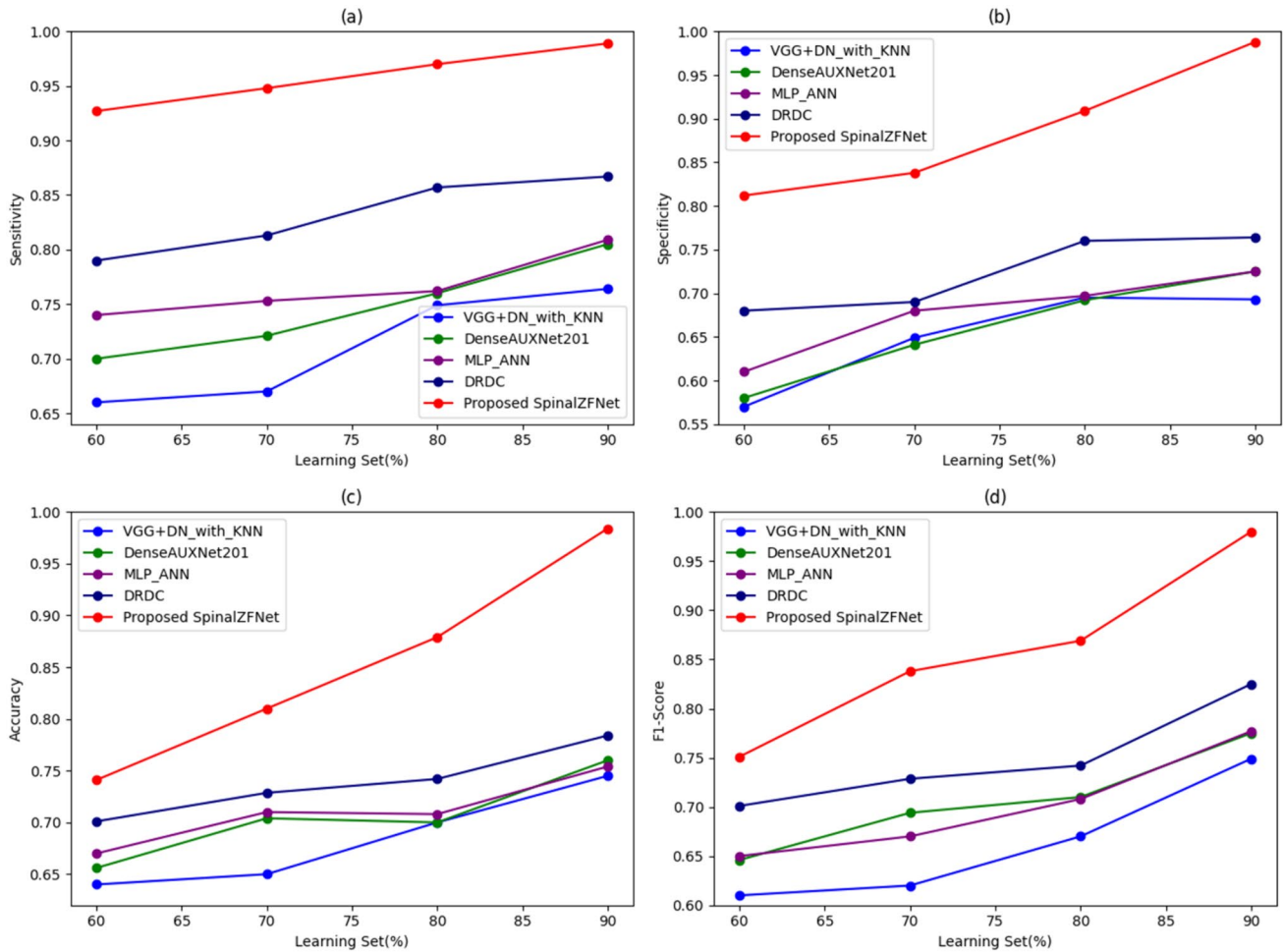
Fig. 8 Comparative analysis of SpinalZFNet using learning set: sensitivity, specificity, accuracy, and F1-score

$$F1 - Score = 2 \times \frac{Precision \times Sensitivity}{Precision + Sensitivity} \tag{38}$$

4.4 Performance Assessment

The *K*-fold value and percentage learning set are used to determine SpinalZFNet's classification performance during kidney disease classification.

Table 2 Comparative analysis of SpinalZFNet with state-of-the-arts models

Variations	Parameters	VGG + DN with KNN	DenseAUXNet201	MLP-ANN	DRDC	SpinalZFNet
Learning set	Sensitivity(%)	76.4	80.5	80.9	86.7	**98.9**
	Specificity(%)	72.3	72.5	72.5	76.4	**98.8**
	Precision(%)	73.3	74.6	74.6	78.6	**97.7**
	Accuracy(%)	74.5	76.0	75.4	78.4	**98.4**
	F1-Score(%)	74.9	77.5	77.7	82.5	**98.0**
K-Fold	Sensitivity(%)	81.7	84.5	85.0	89.4	**99.9**
	Specificity(%)	73.0	79.3	79.4	81.0	**99.6**
	Precision(%)	74.6	80.3	80.4	82.4	**99.6**
	Accuracy(%)	80.5	80.7	82.1	83.5	**99.8**
	F1-Score(%)	78.0	82.4	82.7	85.8	**99.7**

Bold values highlight the highest value

4.4.1 Analysis Using *K*-fold

The performance validation of superiority used for kidney disease classification using *K*-fold is illustrated in Fig. 6. The model's accuracy progression over different epochs is presented in Fig. 6a. It can be observed that with the increase in epochs, there is a consistent improvement in accuracy. For a *K*-fold value 10, the recorded sensitivity for epochs 10 to 80 is 80.9%, 86.3%, 87.1%, 88.0%, and 99.0%, respectively, highlighting the model's learning capability and stability over extended training. In Fig. 6b, the 78.1%, 79.8%, 81.9%, 83.1%, and 85.13% specificity is recorded by SpinalZFNet for *K*-fold values 10 for 10 to 80 epochs. Also, in Fig. 6c the evaluation of performance using accuracy is depicted. For *K*-fold value 10, the SpinalZFNet measured accuracy for epochs 10 to 80 is 79.9%, 81.3%, 83.1%, 85%, and 99%.

4.4.2 Analysis Using Learning Set

The analysis of the performance of SpinalZFNet utilized for the classification of kidney disease using a learning set is shown in Fig. 6. In Fig. 6d, the performance analysis using sensitivity is portrayed. For learning a set of 90%, the SpinalZFNet obtained a sensitivity of 81.2–98.3% for 10–80 epochs. Figure 6e shows that the model achieved a specificity of 75–98.1% on 10–80 epochs. Figure 6f shows

the model's performance using accuracy. The SpinalZFNet's accuracy was measured for epochs 10 with a value of 79.5% and for 80 epochs with a value of 98.4%, respectively.

4.5 Comparative Analysis

Traditional classification models have been used to analyze data, including VGG + DN with KNN [1], DenseAUXNet201 [7], MLP-ANN [18], and DRDC [8]. In this study, SpinalZFNet was compared to these models using the *K*-fold value and percentage of the learning set. Metrics such as sensitivity, specificity, precision, accuracy, F1-score, confusion matrix, and ROC curve were used to evaluate performance.

4.5.1 Analysis Using *K*-fold

The comparative analysis of SpinalZFNet efficiency using the *K*-fold value during kidney disease classification represents in Fig. 7. The performance analysis is based on the sensitivity, specificity, accuracy, and F1-Score of various approaches used for kidney disease classification. The sensitivity measured by SpinalZFNet is 99.9% for the *K*-fold value of 10. The classical classification approaches, like VGG + DN with KNN, DenseAUXNet201, MLP-ANN, DRDC, and proposed SpinalZFNet recorded sensitivity
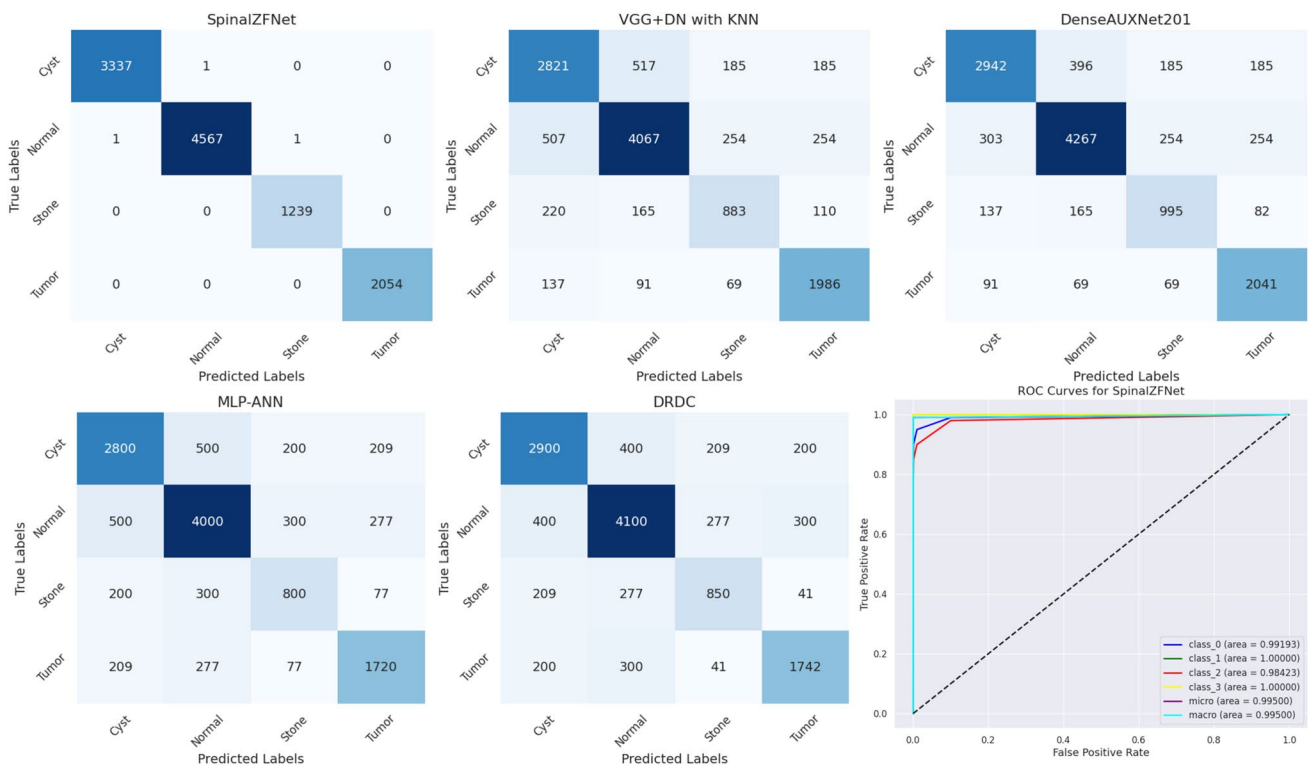

Fig. 9 Performance analysis using confusion matrix of SpinalZFNet with state-of-the-art models and ROC curves for SpinalZFNet

Fig. 10 A sample visualizations of the ZFNet model

of 83.2%, 86.2%, 86.7%, 89.4%, and 99.9%, as shown in Fig. 7. The specificity 99.6% is notably higher in SpinalZFNet, indicating fewer false positives and more reliable performance in distinguishing non-disease cases. Precision is highest for SpinalZFNet at 99.6%, suggesting its predictions are highly trustworthy. The accuracy measured for *K*-fold value of 10 by the disease classification approaches is 80.8% by VGG + DN with KNN, 82.3% by DenseAUXNet201, 83.0% by MLP-ANN, 86.5% by DRDC, and 99.8% by SpinalZFNet. The analysis proved that SpinalZFNet attained better performance than the prevailing DRDC model. The F1-Score for *K*-fold value 10 is 78% by VGG + DN with KNN, 82.4% by DenseAUXNet201, 82.7% by MLP-ANN, 85.8% by DRDC, and 99.7% by SpinalZFNet. The analysis proved that SpinalZFNet performed better than the state-of-the-art models.

4.5.2 Analysis Using Learning Set

The validation of the performance of SpinalZFNet used for kidney disease classification by varying percentages of the learning set is displayed in Fig. 8. Figure 8 presents the evaluation of classification techniques using sensitivity, specificity, precision, accuracy, and F1-Score. The methods used for the classification of kidney disease sensitivity measured by SpinalZFNet, VGG + DN with KNN, DenseAUXNet201, MLP-ANN, and DRDC are 98.9%, 76.4%, 80.56%, 80.97%, and 86.7%, respectively. The specificity recorded for 50% of the learning set by SpinalZFNet is 98.8%. Similarly, the traditional classification techniques, such as VGG + DN with KNN, DenseAUXNet201, MLP-ANN, and DRDC, obtained specificity of 72.3%, 72.5%, 72.5%, and 76.4%. It is proven that the SpinalZFNet achieved the best performance of 97.7% compared to the traditional VGG + DN with KNN, DenseAUXNet201, MLP-ANN, and DRDC approaches, regarding precision, 73.3%, 74.6%, 74.68%, and 78.69% achieved values for classification approaches. Moreover, the analysis of kidney disease classification approaches shows that the accuracy measured by SpinalZFNet for learning set value is 98.4%, and the accuracy measured by prevailing kidney disease classification models, such as VGG + DN with KNN, DenseAUXNet201, MLP-ANN, and DRDC, is 74.5%, 76%, 70.54%, and 78.4%. Lastly, the F1-score of VGG + DN with KNN is 74.9%, DenseAUXNet201 77.5%, MLP-ANN 77.7%, DRDC 82.5%, and SpinalZFNet 98.0%. The analysis revealed that the SpinalZFNet performed better than the prevailing state-of-the-art models used for disease classification.

5 Comparative Discussion

The classification performance of SpinalZFNet is determined by validating the performance with prevailing techniques used for the classification of kidney disease. The results obtained by the classification approaches for different performance metrics validated using a learning set, and *K*-fold are given in Table 2. Based on experimental investigation, it is proven that the SpinalZFNet, with the learning set

Table 3 Performance comparative analysis of proposed SpinalZFNet with state-of-the-arts models

Architecture	Parameters (millions)	FLOPs (billions)	Training time (h)	Inference time (ms)
VGG + DN with KNN	138	19.6	12	50
DenseAUXNet201	18	4.0	10	30
MLP-ANN	60	12.0	15	45
DRDC	45	9.0	13	35
SpinalZFNet	23	5.2	9	28
SpinalNet	**5**	**1.5**	**8**	**25**

Bold values highlight the highest value

evaluation, emerges as the most competent model in terms of sensitivity, specificity, precision, accuracy, and F1-Score at 98.9%, 98.8%, 97.7%, 98.4%, and 98.0%. With K-Fold evaluation, the SpinalZFNet maintains the best performance with a maximum sensitivity of 99.9%, specificity of 99.6%, precision of 99.6%, accuracy of 99.8%, and F1-score of 99.7%. In contrast, during kidney disease classification, the SpinalZFNet effectively solves complexity issues using its pre-trained architecture. Similarly, it utilizes highly efficient feature maps to enhance classification accuracy by eliminating the undesired features that lead to incorrect predictions. The comparison of the provided confusion matrices for the models MLP-ANN, DRDC, SpinalZFNet, VGG + DN with KNN, and DenseAUXNet201 yields several insights into their performance in classifying the four classes: cyst, normal, stone, and tumor. SpinalZFNet stands out as the best model, particularly for the critical "Tumor" classification, where it has the highest number of true positives and lowest false negatives, making it the most suitable model among those compared for classifying the CT images of Cyst, Normal, Stone, and Tumor. SpinalZFNet model performance comparison with all models using a confusion matrix provides in Fig. 9. The ROC curve illustrates the performance of the SpinalZFNet, a classifier handling multiple classes. The curves for individual classes: class 0 cyst, class 1 normal, class 2 stone, and class 3 tumor, all show high AUC values, with classes 1 and 3 achieving perfection with an AUC of 1.00%. This indicates an exceptional discriminative power where true positives are maximized, and false positives are minimized or nonexistent, especially for classes 1 and 3. Additionally, the micro-average and macro-average ROC curves indicate a high overall performance across classes, with both curves achieving AUCs around 0.995%, signifying the model's consistent and robust classification ability. The dashed line represents the baseline of a random chance classifier, which all the individual ROC curves significantly outperform, demonstrating the classifier's effectiveness.

The sample visualization of the ZFNet model process represents in Fig. 10. The following figure shows the process of training the ZFNet model and categorizing the images based on extracted features, which are highlighted by the model to explain how the procedure helps to classify the disease necessary for classifying the images in training.

The performance comparison of the proposed SpinalZFNet with state-of-the-art models is provided in Table 3. The SpinalZFNet achieves this superior performance with significantly fewer parameters and lower computational costs than other state-of-the-art models. However, Spinal-Net demonstrates even better efficiency than the proposed model SpinalZFNet.

6 Conclusion

This research introduces a hybrid deep learning model, SpinalZFNet, to classify kidney disease accurately from CT images. The median filter is initially used to pre-process the acquired input CT images, and the pre-processed image is segmented using ENet. The segmented image further allows for image augmentation, and the different features are extracted from the augmented image using feature extractors. Finally, the SpinalZFNet is used to classify kidney disease into four distinct categories: normal, cyst, tumor, and stone. Moreover, we evaluated the model's performance using preferred quality metrics. Our SpinalZFNet outperforms the existing models and achieved 99.9% sensitivity, 99.6% specificity, 99.8% accuracy, and 99.7% F1-Score. Our experimental results prove that the SpinalZFNet can accurately classify four different classifications of kidney disease using CT images.

Author contributions Faiqa Maqsood: conceptualization, methodology, formal analysis, Faiqa Maqsood and Muhammad Mumtaz Ali: the original draft, writing, review and editing, Wang Zhenfei, Raheem Sarwar, and Baozhi Qiu: supervision. all authors read, gave suggestions, and agreed to the published version of the manuscript.

Data availability The publicly available dataset used in this manuscript.

Declarations

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

References

1. Rajinikanth V, Vincent PMDR, Srinivasan K et al (2023) A framework to distinguish healthy/cancer renal CT images using the fused deep features. Front Public Health 11:1109236. https://doi.org/10.3389/fpubh.2023.1109236

2. Kovesdy CP (2022) Epidemiology of chronic kidney disease: an update 2022. Kidney Int Suppl 12:7–11. https://doi.org/10.1016/j.kisu.2021.11.003

3. Bhandari M, Yogarajah P, Kavitha MS et al (2023) Exploring the capabilities of a lightweight CNN model in accurately identifying renal abnormalities: cysts, stones, and tumors, using LIME and SHAP. Appl Sci 13:3125. https://doi.org/10.3390/app13053125

4. Wu Yu, Yi Z (2020) Automated detection of kidney abnormalities using multi-feature fusion convolutional neural networks. Knowl-Based Syst 200:105873. https://doi.org/10.1016/j.knosys.2020.105873

5. Islam MN, Hasan M, Hossain MK et al (2022) Vision transformer and explainable transfer learning models for auto-detection of kidney cyst, stone and tumor from CT-radiography. Sci Reports 12:11440. https://doi.org/10.1038/s41598-022-15634-4

6. Sudharson S, Kokil P (2021) Computer-aided diagnosis system for the classification of multi-class kidney abnormalities in the noisy ultrasound images. Comput Methods Programs Biomed 205:106071. https://doi.org/10.1016/j.cmpb.2021.106071

7. Mahmud S, Abbas TO, Mushtak A et al (2023) Kidney cancer diagnosis and surgery selection by machine learning from CT scans combined with clinical metadata. Cancers 15:3189. https://doi.org/10.3390/cancers15123189

8. Badawy M, Almars AM, Balaha HM et al (2023) A two-stage renal disease classification based on transfer learning with hyperparameters optimization. Front Med 10:1106717. https://doi.org/10.3389/fmed.2023.1106717

9. Raji CG, Anand, et al (2017) Computer-based prognosis model with dimensionality reduction and validation of attributes for prolonged survival prediction. Inform Med Unlocked 9:93–106. https://doi.org/10.1016/j.imu.2017.07.002

10. Gharaibeh M, Alzubi D, Abdullah M et al (2022) Radiology imaging scans for early diagnosis of kidney tumors: a review of data analytics-based machine learning and deep learning approaches. Big Data Cog Comp 6:29. https://doi.org/10.3390/bdcc6010029

11. Chen G, Ding C, Li Y et al (2020) Prediction of chronic kidney disease using adaptive hybridized deep convolutional neural network on the internet of medical things platform. IEEE Access 8:100497–100508. https://doi.org/10.1109/ACCESS.2020.2995310

12. Suarez-Ibarrola R, Hein S, Reis G et al (2020) Current and future applications of machine and deep learning in urology: a review of the literature on urolithiasis, renal cell carcinoma, and bladder and prostate cancer. World J Urology 38:2329–2347. https://doi.org/10.1007/s00345-019-03000-5

13. Yu K-H, Beam AL, Kohane IS (2018) Artificial intelligence in healthcare. Nat Biomed Engine 2:719–731. https://doi.org/10.1038/s41551-018-0305-z

14. Kijowski R, Liu F, Caliva F et al (2020) Deep learning for lesion detection, progression, and prediction of musculoskeletal disease. J Magn Res Imaging 52:1607–1619. https://doi.org/10.1002/jmri.27001

15. Patro KK, Jaya Prakash A, Jayamanmadha Rao M et al (2022) An efficient optimized feature selection with machine learning approach for ECG biometric recognition. IETE J Res 68:2743–2754. https://doi.org/10.1080/03772063.2020.1725663

16. Patro KK, Allam JP, Neelapu BC et al (2023) Application of kronecker convolutions in deep learning technique for automated detection of kidney stones with coronal CT images. Info Sci 640:119005. https://doi.org/10.1016/j.ins.2023.119005

17. Tabibu S, Vinod PK, Jawahar CV (2019) Pan-renal cell carcinoma classification and survival prediction from histopathology images using deep learning. Sci Rep 9:10509. https://doi.org/10.1038/s41598-019-46718-3

18. Cai J, Liu M, Zhang Q et al (2022) Renal cancer detection: fusing deep and texture features from histopathology images. BioMed Res Int 2022:e9821773. https://doi.org/10.1155/2022/9821773

19. Shehata M, Alksas A, Abouelkheir RT et al (2021) A comprehensive computer-assisted diagnosis system for early assessment of renal cancer tumors. Sensors 21:4928. https://doi.org/10.3390/s21144928

20. Maheshan CM, Prasanna Kumar H (2019) Performance of image pre-processing filters for noise removal in transformer oil images at different temperatures. SN Appl Sci 2:67. https://doi.org/10.1007/s42452-019-1800-x

21. Paszke A, Chaurasia A, Kim S et al (2016) ENet: a deep neural network architecture for real-time semantic segmentation. arXiv. https://doi.org/10.48550/arXiv.1606.02147

22. Sharma A, Yadav DP, Garg H et al (2021) Bone cancer detection using feature extraction based machine learning model. Comput Math Methods Med 2021:e7433186. https://doi.org/10.1155/2021/7433186

23. Murillo AC, Guerrero JJ, Sagues C (2007) SURF features for efficient robot localization with omnidirectional images. In: Proceedings 2007 IEEE International Conference on Robotics and Automation. pp 3901–3907. https://doi.org/10.1109/ROBOT.2007.364077

24. Zhang D, Li Q, Yang G et al (2017) Detection of image seam carving by using weber local descriptor and local binary patterns. J Inf Secur Appl 36:135–144. https://doi.org/10.1016/j.jisa.2017.09.003

25. Chilakala LR, Kishore GN (2021) Optimal deep belief network with opposition-based hybrid grasshopper and honeybee optimization algorithm for lung cancer classification: a DBNGHHB approach. Int J Imaging Syst Technol 31:1404–1423. https://doi.org/10.1002/ima.22515

26. Antonik P, Marsal N, Brunner D et al (2019) Human action recognition with a large-scale brain-inspired photonic computer. Nat Mach Intell 1:530–537. https://doi.org/10.1038/s42256-019-0110-8

27. Kabir HMD, Abdar M, Khosravi A, Jalali SMJ et al (2023) SpinalNet: deep neural network with gradual input. IEEE Trans Artif Intell 4:1165–1177. https://doi.org/10.1109/TAI.2022.3185179

28. Fu L, Feng Y, Majeed Y et al (2018) Kiwifruit detection in field images using faster R-CNN with ZFNet. IFAC-Pap 51:45–50. https://doi.org/10.1016/j.ifacol.2018.08.059

29. Chen J, Shan S, He C et al (2010) WLD: a robust local image descriptor. IEEE Trans Pattern Anal Mach Intell 32:1705–1720. https://doi.org/10.1109/TPAMI.2009.155

30. Kasim A, Harjoko A (2017) Batik classification with artificial neural network based on texture-shape feature of main ornament. Int J Intell Syst Appl 9:55–65. https://doi.org/10.5815/ijisa.2017.06.06

31. Al Khalil Y, Amirrajab S, Lorenz C et al (2023) Reducing segmentation failures in cardiac MRI via late feature fusion and GAN-based augmentation. Comput Biol Med 161:106973. https://doi.org/10.1016/j.compbiomed.2023.106973

32. Nguyen H-T, Li S, Cheah CC (2022) A layer-wise theoretical framework for deep learning of convolutional neural networks. IEEE Access 10:14270–14287. https://doi.org/10.1109/ACCESS.2022.3147869

33. Raji CG, Vinod Chandra SS (2016) Prediction and survival analysis of patients after liver transplantation using rbf networks. Data Mining and Big Data. Springer, Cham, pp 147–155

34. Raji CG, Chandra SSV (2017) Various medical aspects of liver transplantation and its survival prediction using machine learning techniques. Indian J Sci Technol 10(13):1–17. https://doi.org/10.17485/ijst/2017/v10i13/94111

Authors and Affiliations

Faiqa Maqsood[1] · Wang Zhenfei[1] · Muhammad Mumtaz Ali[1] · Baozhi Qiu[1] · Naveed Ur Rehman[1] · Fahad Sabah[2] · Tahir Mahmood[3] · Irfanud Din[4] · Raheem Sarwar[5]

✉ Raheem Sarwar
R.Sarwar@mmu.ac.uk

Faiqa Maqsood
faiqamaqsood@gs.zzu.edu.cn

Wang Zhenfei
iezfwang@zzu.edu.cn

Muhammad Mumtaz Ali
muali@gs.zzu.edu.cn

Baozhi Qiu
iebzqiu@zzu.edu.cn

Naveed Ur Rehman
naveed@gs.zzu.edu.cn

Fahad Sabah
fahad.sabah@emails.bjut.edu.cn

Tahir Mahmood
tahirmahmood@dongguk.edu

Irfanud Din
irfan@newuu.uz

[1] School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou 450001, China

[2] Beijing University of Technology, Beijing 100124, China

[3] Division of Electronics and Electrical Engineering, Dongguk University, Seoul 04620, South Korea

[4] Department of Computer Science, New Uzbekistan University, Tashkent 100174, Uzbekistan

[5] OTEHM, Faculty of Business and Law, Manchester Metropolitan University, M15 6BH Manchester, UK