Global Health: Journal of Health Sciences, Public Health and Pharmacy

E-ISSN: 3062-8687
P-ISSN: 3063-7465

(Research) Article

CKD Detection Using CNN on Ultrasound Images Based on Estimated Glomerular Filtration Rate (EGFR) Values

Akhmad Subarkah[1]*, Edy Susanto[2], and Agung Nugroho Setiawan[3]

[1-3] Poltekes Kemenkes Semarang
* Corresponding Author : akhmad.subarkah@siemens-healthineers.com

Abstract: Chronic kidney disease (CKD) is a progressive and irreversible decline in kidney function that, if left untreated, can lead to serious complications. Ultrasonography (USG) is a widely used imaging modality for detecting CKD, yet its interpretation remains highly dependent on the radiologist's expertise. This study aims to develop a CKD detection system using a convolutional neural network (CNN) on kidney ultrasound images based on estimated glomerular filtration rate (eGFR), and to evaluate the system's performance. This research employed a research and development (R&D) approach with an experimental design. The dataset consisted of kidney ultrasound images from CKD and non-CKD patients with corresponding eGFR values. The methodology included image preprocessing, CNN model training, and accuracy evaluation using classification metrics. The results demonstrated that the developed CNN model achieved a total accuracy of 97% on internal test data and 95.8% on external validation. The model's sensitivity reached 100% for the normal category, 91.67% for CKD stage 4, and 90% for CKD stage 5. Specificity exceeded 96% across all categories, with high precision and F1-scores above 94% for all classes. This system has proven to be effective as a diagnostic support tool for automatically detecting CKD through kidney ultrasound imaging. Its advantages lie not only in accurately classifying CKD from USG images but also in correlating the classification results with patients' eGFR values. This provides more precise clinical information and supports appropriate CKD staging and management planning.

Keywords: Chronic Kidney Disease (CKD); Convolutional Neural Network (CNN); Estimated Glomerular Filtration Rate (eGFR); Medical Image Classification**;** Ultrasonography.

Received: June 29, 2025
Revised: August 15, 2025
Accepted: August 30, 2025
Published: September 30, 2025
Curr. Ver.: September 30, 2025

1. Introduction

Chronic kidney disease (CKD) is a kidney disease that can lead to end-stage renal failure if not treated properly. Chronic kidney disease (CKD) is a pathophysiological process with numerous aetiologies that cause progressive and irreversible decline in kidney function to the point where the body fails to support metabolic function and maintain adequate fluid and electrolyte balance, leading to uraemia (1).

The 2018 Indonesian Basic Health Research Data shows that the number of chronic kidney disease patients has reached 713,783 people. The prevalence of chronic kidney disease among Indonesians aged 15 years and above diagnosed by a doctor is 0.38%. In East Java, the prevalence is 0.29% (2). Specifically, a meta-analysis found that the average prevalence of stage 5 chronic kidney disease (CKD) was 13.4%, stage 3-4 was 10.6%, stage 1 was 3.5%, and stage 2 was 3.9%. Overall, the global prevalence of CKD is estimated to be 11-13%.

Early detection of CKD is crucial to prevent more serious complications. Radiology is one of the medical supports that is very helpful in establishing a diagnosis of CKD. Image-based assessment has always been the conventional diagnostic method in clinical practice. Magnetic resonance imaging (MRI), computed tomography (CT), ultrasound, X-rays, histology, and positron emission tomography are just a few examples of the various types of medical imaging, all of which have unique applications in the medical field.

Ultrasonography is one of the conventional imaging modalities in radiology. Ultrasonography is a non-invasive imaging modality that does not involve ionising radiation and generally provides good diagnostic results. The ultrasonography technique involves the use of sound waves at frequencies higher than those detectable by the human ear, in the range of 1–20 megahertz. Various abdominal organs such as the liver, gallbladder, bile ducts, pancreas, spleen, kidneys, and retroperitoneal structures can be visualised anatomically and pathologically through this examination. This technique essentially combines advantages such as non-invasiveness, ease of implementation, and accuracy (3). However, the interpretation of ultrasound images by radiologists still requires sufficient training and experience.

Chronic Kidney Disease is generally diagnosed by calculating the glomerular filtration rate. However, the definition of CKD in terms of structural changes makes it possible to diagnose CKD in patients whose estimated glomerular filtration rate is still normal. Imaging of the kidney structure, more specifically ultrasound (USG), helps to observe kidney defects, which are known symptoms of CKD. Renal imaging using USG not only provides a good diagnosis for CKD cases but also aids in treatment. The advantages of USG in diagnosing the cause of kidney disease are that it is non-invasive, universally available, and does not utilise radiation. USG can reveal the shape and size, relative echogenicity of the parenchyma, spatial dimensions of the collecting system and ureter, and features outside the normal renal line such as masses or vascular lesions.

The potential of ultrasound evaluation for kidney disease includes the shape and size of the kidneys, renal parenchymal echogenicity, urinary spaces, masses and renal blood vessels. Common features in patients with chronic kidney disease (CKD) are usually reduced kidney size, cortical thinning and cystic masses with increased cortical echogenicity in some CKD patients. Ultrasound can be used to investigate the aetiology of CKD, such as obstructive uropathy, polycystic kidney disease, reflux nephropathy, or interstitial nephritis, among others. A linear correlation between ultrasound-measured renal parenchymal thickness and kidney length with GFR levels was observed by Christy et al. (2015), where statistical significance ($p < 0.001$) was demonstrated, thereby establishing ultrasound as a reliable modality for diagnosing CKD. The aim of this study was to identify the radiological imaging characteristics of ultrasound in CKD patients; all CKD patients were taken from the medical records of Sanglah General Hospital during the period of 2015. Information obtained from ultrasound examination results in CKD patients can facilitate the diagnosis of this disease so that timely management can be carried out (4).

Based on preliminary studies at Hermina Hospital in Depok, B-Mode ultrasound is still used in the examination or diagnosis of CKD in patients, namely to measure kidney length, cortex thickness and renal parenchyma echogenicity. The development of artificial intelligence in ultrasonography has progressed rapidly, one example being the use of Convolutional Neural Networks (CNN). CNN is a deep learning model that has been proven effective in object recognition and visual data clustering, which can assist in identifying the results of an ultrasonography examination. CNN can be used to detect specific features in ultrasonography images related to CKD (5).

Currently, advances in deep learning technology, particularly convolutional neural network (CNN) algorithms, have proven effective in medical image pattern recognition, including the detection of kidney characteristics from ultrasound results. CNNs are capable of automatically recognising echogenicity patterns, kidney size, and cortical thickness through visual feature extraction. One important aspect that can be optimised in this process is the correlation between the results of kidney ultrasound image analysis and the patient's estimated glomerular filtration rate (eGFR) value. The eGFR value is a key parameter in determining the stage of chronic kidney disease (CKD) and is clinically used as a standard for diagnosis and management.

The correlation between renal ultrasound images and eGFR values has been proven in various studies, in which increased renal echogenicity and decreased cortical thickness on ultrasound are linearly related to decreased eGFR. Therefore, the development of a CNN-based CKD detection system linked to eGFR values not only improves the objectivity of image analysis results but also supports the accuracy of disease stage stratification. This system is expected to be able to classify the patient's kidney condition based on ultrasound morphological images while identifying the eGFR category, so that CKD diagnosis can be made more quickly, accurately, and in accordance with clinical standards.

The objective of this thesis is to develop a CKD detection system using CNN on ultrasound images based on glomerular filtration rate (GFR) values. The methods used include collecting ultrasound image datasets from CKD and non-CKD patients, preprocessing images to improve image quality, training CNN with the prepared dataset, validating model performance using relevant metrics, and analysing the results to determine the accuracy of CKD detection. To date, the interpretation of renal ultrasound is still highly dependent on radiologists, and there is no integrated AI system available in hospitals as an assistant for the automatic early detection of CKD.

2. Preliminaries or Related Work or Literature Review

Kidney

The human kidneys are located in the posterior part of the body in the upper retroperitoneal space, precisely on both sides of the aorta and inferior vena cava between the twelfth thoracic vertebra and the third lumbar vertebra. Typically, each adult kidney weighs approximately 150 grams and measures 11–14 cm in length. The right kidney is several centimetres smaller than the left kidney because the liver is located beneath it (6). For protection against external trauma, all kidneys are enclosed in a strong fibrous capsule. On the medial side of the kidney is the hilum, which forms several concave areas for the renal arteries and veins, lymphatics, nerves, and ureter to pass through and enter the kidney, transporting waste products formed in the kidney to the bladder (7).

The kidney has two parts, the outer part known as the cortex and the inner part known as the medulla. The medulla has 8 to 10 renal pyramids, which are cone-shaped masses of tissue. These pyramids begin at the junction of the cortex and medulla and end at the papilla. The tip of each renal pyramid forms an open-ended minor calyx that extends downward as a major calyx to collect urine. Within the renal cortex are millions of nephrons, which are the functional units of the kidney consisting of a glomerulus, proximal convoluted tubule, Henle's loop, and distal convoluted tubule.

Image description:

1. Capsula
2. Cortex
3. Renal Artery
4. Renal Vein
5. Renal Pelvis
6. Perirenal Fat
7. Ureter
8. Medula
9. Renal Colums
10. Papilae
11. Renal Pyramid
12. Corticmedulary Juncion
13. Calyces
14. Arcutary Artery

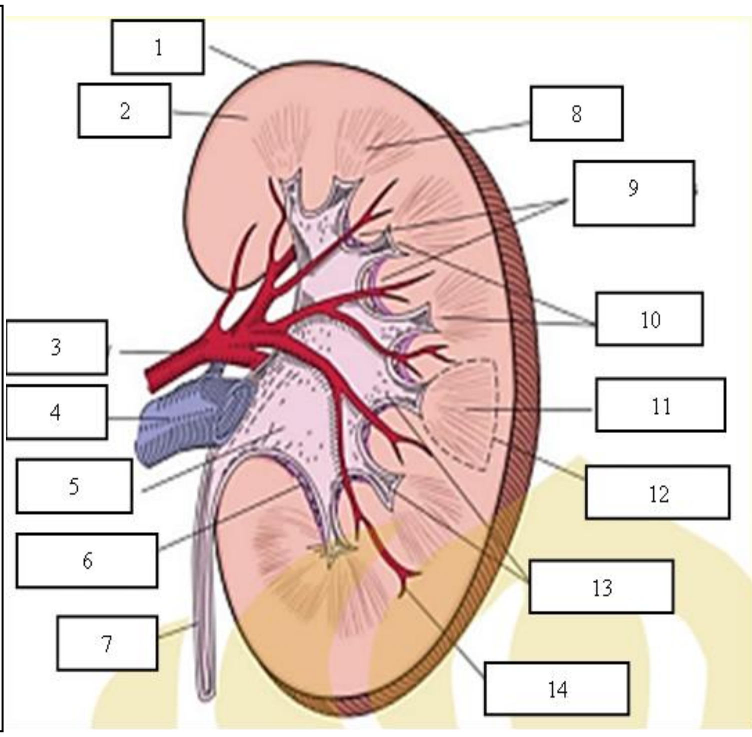

Figure 1. Anatomy of the Kidney *(8)*

The kidneys play an important role in the further metabolism of ammonia, urea, and creatinine produced in proteins, uric acid, drugs, and toxins. In addition, homeostasis, the proper balance of fluid and electrolyte regulation, is necessary for the kidneys. The kidneys work together with excretion to control the levels of bicarbonate, sodium, calcium, chloride, and potassium in the body. Regulatory activities involve blood pressure, acidity, calcium in the body, and phosphate levels. This helps maintain homeostasis in the body. This is a very important function in the body because, without it, the body would be at risk of blood pressure fluctuations (6).

CKD (*Chronic kidney disease*)

CKD is defined as a structural or functional abnormality of the kidneys, lasting more than 3 months, with implications for health, and classified from 1 to 5 according to GFR or albuminuria, with or without markers of kidney damage. CKD is defined as a functional or structural abnormality of the kidneys lasting for 3 months, with or without a decrease in eGFR, manifesting as urological disease/symptoms with abnormalities found in blood and urine tests or abnormalities detected in imaging tests or with an eGFR of less than 60 ml per minute per 1.73 m^2 for more than 3 months, with or without kidney damage (9).

Factors believed to be responsible for the development of chronic kidney disease include albuminuria, hypertension, hyperglycaemia, and dyslipidaemia. The occurrence of glomerular and tubulointerstitial sclerosis and fibrosis varies considerably among subjects (20). In the earliest stages of chronic kidney disease, there is a loss of renal reserve, with basal GFR remaining normal or even increasing. Progressively, albeit slowly, nephron function deteriorates, as reflected in serum urea and creatinine levels. With a GFR decline of up to 60%, some patients remain asymptomatic and do not complain of any symptoms, and only then do creatinine and urea levels begin to rise. Symptoms begin to develop in patients. Patients present with a GFR of 30%, nocturia, weakness, nausea, decreased appetite, and weight loss. Below 30%, patients experience anaemia, high blood pressure, and phosphorus-calcium metabolism disorders. Itching, nausea, and vomiting occur due to high phosphorus levels. This makes patients more susceptible to infections such as urinary tract infections or gastrointestinal infections. Water balance disorders that occur are hypervolaemia or hypovolaemia, and electrolyte imbalances include sodium and potassium. At a GFR of less than 15%, more serious symptoms and complications arise, and patients will require renal replacement therapy. At this point, patients are said to have reached the stage of renal failure (10).

Teknik Scaning Ultrasonografi Ginjal

Ultrasonography has high value in the medical assessment of renal pathology. Because it can describe the structure and function of the kidneys quite clearly, ultrasonography can be used not only for early detection of renal pathology but also for routine follow-up of diagnosed cases (11). This method can be further developed through additional research to improve its diagnostic accuracy, combining it with renal artery RI with colour Doppler and other modalities. Renal ultrasound is a non-invasive method that determines the morphology and function of the kidneys. This method produces images of the kidneys using sound waves of varying frequencies to record pathological conditions or medical situations developing in or around the kidneys. It works by emitting high-frequency sound waves (1-10 MHz) through a transducer placed on the skin in the kidney area. The reflections from the kidney tissue process the information for imaging.

Typically, the kidneys measure between 95-110 mm in length and 50-60 mm in width. Usually, the kidneys are labelled as such if they are larger or exceed this scale; if they are smaller, this indicates (11). Increased renal echogenicity generally indicates pathology in the parenchyma. The echostructure can manage this. The cortex-medulla boundary, a clear demarcation between the cortex and medulla, provides insight into kidney health. If the boundary is blurred, it suggests an inflammatory or destructive condition (12). The pyelocaliceal system, pathologies that can be diagnosed with the pyelocaliceal system include hydronephrosis, where there is obstructive enlargement or reflux (13).

The ultrasound examination technique is performed with the patient lying supine on the examination table or in the left lateral decubitus position. Meanwhile, the left kidney can be imaged in the right lateral decubitus position. The scanning technique is performed longitudinally and transversely using the acoustic window of the liver on the right kidney and on the left kidney using the acoustic window on the spleen. Normal kidney images show a length of at least 9 to 12 cm, a width of 2 to 6 cm, low echogenic cortex affecting the liver, a hypoechoic pelviocalyx system without dilation, hyperechoic sinuses, and hypoechoic medulla.

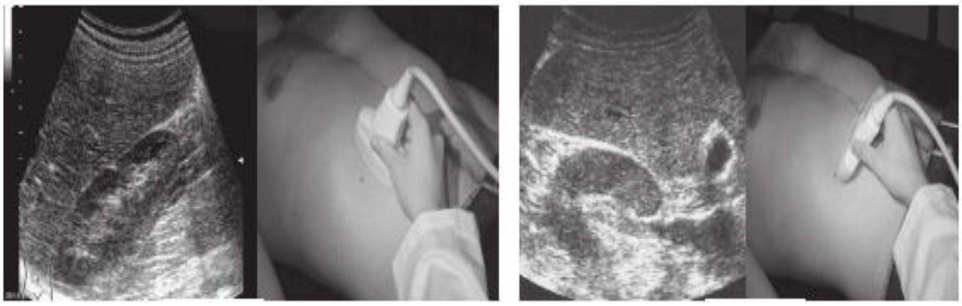

Figure 2. (a) Longitudinal screening technique and results and (b) Transverse screening technique and results (14)

Kidney ultrasound examinations are generally performed to diagnose chronic kidney disease (CKD) and kidney failure, as well as to assess abnormalities such as kidney stones, cysts, or masses. In addition, this examination also aims to check urine flow and detect any obstructions that may interfere with kidney function. Kidney ultrasound is also often used to monitor the progression of kidney disease on a regular basis, making it easier for doctors to determine the appropriate treatment (13).

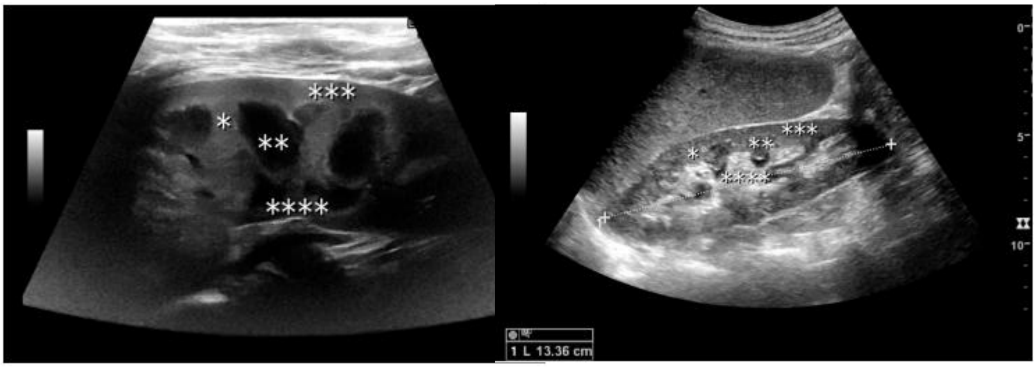

Figure 3. Normal Adult Kidney (15)

The image shows the results of a 2D ultrasound of a normal adult kidney, with clear kidney structures. The renal cortex appears homogeneous with lower echogenicity than the renal sinus, which appears hyperechoic due to its fat and blood vessel content. The renal pelvis appears centred with no signs of dilation or obstruction. The size and shape of the kidneys are symmetrical, indicating no anatomical or pathological abnormalities(15).

Figure 4. Chronic kidney disease (15)

Ultrasound image of the kidney caused by glomerulonephritis with increased echogenicity and decreased cortical thickness. The measurement of kidney length on the ultrasound image is illustrated with a "+" sign and a dotted line (15).

Deep learning Convolutional Neural Network (CNN)

Deep learning is a branch of machine learning, namely a neural network system with three or more layers. This neural network was created to mimic the human brain with its learning abilities, supported by large amounts of data. Although a single-layer neural network is still capable of making predictions, the more hidden layers there are, the more it will help optimise performance and improve accuracy. Deep learning provides powerful developments for Artificial Intelligence (AI) applications and services that lead to automation, providing analytical and physical actions without human assistance. Deep learning technology is behind everyday products and services (such as digital assistants, voice-activated remote TVs, and credit card fraud detection) as well as future products and services (such as self-driving cars).

Deep learning enables computational models to have multiple layers of abstraction. This has led to improved research in various fields, including but not limited to speech recognition and object detection or recognition in photographs. The back-propagation algorithm views deep learning as learning to discover complex structural patterns. The weight of each parameter must be changed by the amount of weight in each model parameter to obtain a more abstract hierarchical level of representation by adjusting the low-level data so that it is more consistent with the data (16).

Convolutional Neural Network (CNN) is the most widely used type of neural network for image recognition. CNN can recognise objects in images. CNN is an approach based on the method used to manifest visual perception in most mammalian species. The CNN algorithm is very precise. CNN can also be used to analyse objects in image data. It has been mentioned that this algorithm is the best model for object recognition. CNN techniques are not significantly different from other artificial neural network techniques such as backpropagation or perceptron because CNNs also consist of neurons with weights, biases, and activation functions similar to those in other Artificial Neural Networks (ANNs) (17).

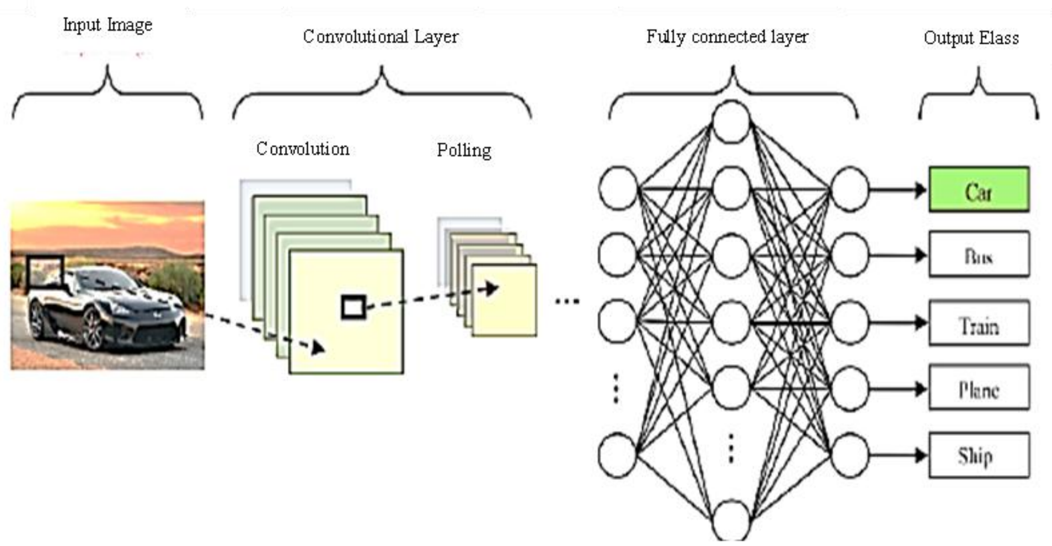

Figure 5. Example of a CNN Algorithm (18)

The architecture of CNNs is similar to that of artificial neural networks (ANNs). Just like human neural networks, CNNs also consist of simple and complex layers, which are designed alternately like the layers in the human brain. CNNs basically consist of convolution layers and pooling layers (19).

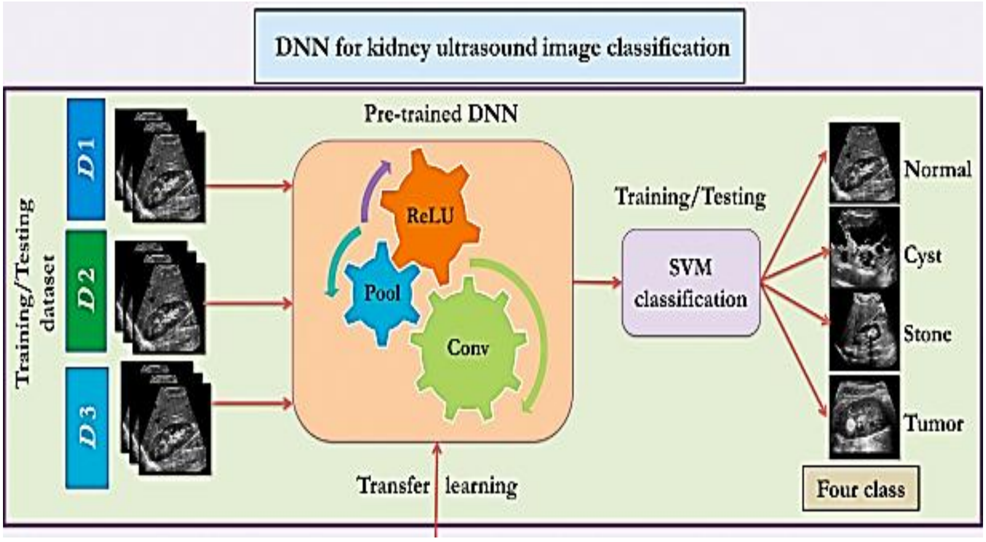

Figure 6. Ultrasound Image CNN Process (20)

This illustration shows the process of classifying kidney ultrasound images using Deep Neural Network (DNN) and Support Vector Machine (SVM) approaches. Data from three datasets (D1, D2, and D3) were used as input to train and test the model. The process begins with a DNN that includes convolution (Conv), pooling (Pool), and ReLU activation functions to extract important features from renal ultrasound images. Through a transfer learning approach, the features extracted by the DNN are then used by the SVM to classify images into four categories: normal, cyst, kidney stone, and tumour. This combination of DNN and SVM enables the system to work efficiently in analysing and identifying abnormalities in renal ultrasound images with a high degree of accuracy.

3. Proposed Method

This type of research is development research using the Research and Development (R&D) method adapted from Sugiyono's (2019) development research model, with an experimental approach (21). Development research was chosen because it aims not only to understand a phenomenon but also to produce a product in the form of an artificial intelligence-based classification system that can be applied as a diagnostic tool. In this case, the system developed is a method for detecting Chronic Kidney Disease (CKD) using a Convolutional Neural Network (CNN) on kidney ultrasound images.

A quantitative approach was used because this study measured system performance through statistical figures, such as accuracy, sensitivity, and specificity. The test results were processed numerically to determine how well CNN was able to detect and classify CKD based on ultrasound images (22).

The research model used consists of five main stages of system development, namely planning, system creation, testing, validation, and evaluation. Each stage is arranged sequentially and is iterative in nature, using the principles of agile development. This means that if there are deficiencies or inconsistencies in one stage, improvements can be made immediately before proceeding to the next stage.

The CNN system designed in this study is tasked with classifying kidney ultrasound images into three categories, namely Normal, CKD Stage 4, and CKD Stage 5. The CNN model was chosen because it has the ability to automatically recognise complex visual patterns without requiring manual feature extraction, making it very suitable for classifying medical images such as ultrasounds.

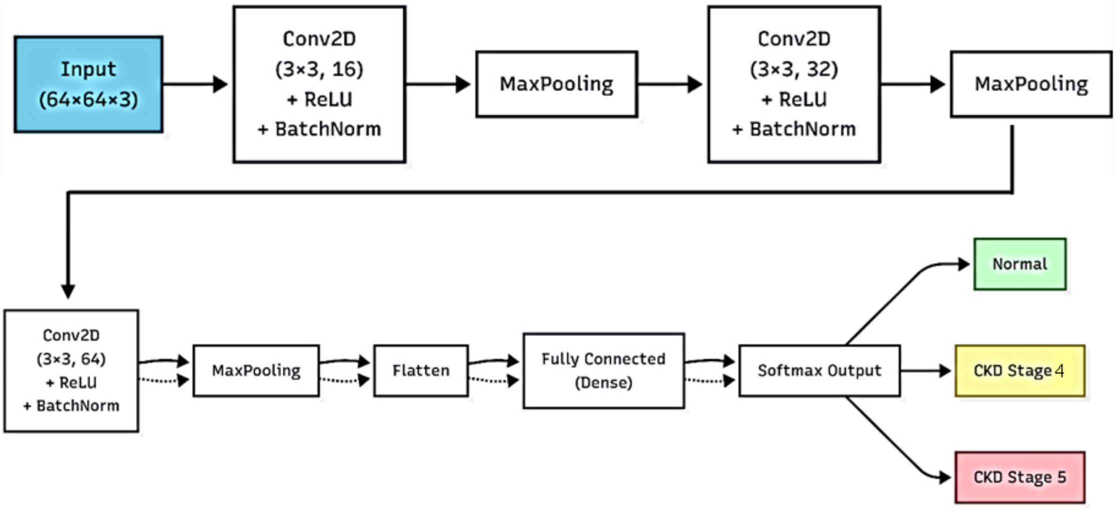

Figure 7. CNN model architecture flowchart

4. Results

The Ability of CNN Models to Classify CKD Based on Ultrasound Images

CNN Model Training Process

The CNN training model was implemented using a comprehensively curated dataset consisting of kidney ultrasound images that had been medically classified into three main diagnostic categories: Normal, CKD Stage 4, and CKD Stage 5. The training process was designed with optimal configuration to achieve maximum performance in recognising complex pathological patterns in kidney ultrasound images.

The training parameter configuration was determined based on experimental results to achieve an optimal balance between accuracy and computational efficiency. The number of epochs was set in the range of 30-50 with the implementation of an early stopping mechanism to prevent overfitting, which can reduce the model's generalisation ability. The early stopping technique monitors the model's performance on the validation set and stops the training process when there is no significant improvement in several consecutive epochs.

A batch size of 32 was selected as the optimal configuration that provides a balance between gradient stability and memory usage efficiency. This batch size allows the model to learn from sufficient variation in each iteration while still being able to run on hardware with standard specifications. A validation split of 20% of the total training data is used to monitor model performance in real time during the training process, enabling early detection of potential overfitting or underfitting.

The data augmentation techniques applied include rotation within a range of ±15°, horizontal and vertical translation of ±10%, and zoom with a variation of ±20%. This augmentation strategy is designed to improve the robustness of the model against variations in the orientation and position of the ultrasound probe that often occur in clinical practice, while also increasing the size of the effective dataset without requiring additional data acquisition.

Classification and Characteristic Recognition Capabilities

The system demonstrates highly specific classification capabilities in recognising different renal morphological characteristics for each diagnostic category. The pattern recognition capabilities developed enable the system to identify complex and often subtle diagnostic features in ultrasound images.

For normal case classification, the system is capable of recognising homogeneous renal echogenicity within the normal range, which is a key indicator of healthy kidney function. A clear and distinct corticomedullary boundary can be detected with high accuracy, indicating good differentiation between the renal cortex and medulla. The system can also analyse kidney size and ensure it is within the normal range according to anthropometric standards. The absence of signs of atrophy or fibrosis is an important parameter that the system can identify to ensure normal classification.

In cases of Stage 4 CKD, the system demonstrates superior ability to detect increased renal cortical echogenicity, which is an early manifestation of glomerular fibrosis and sclerosis. Significant reduction in cortical thickness can be measured and analysed quantitatively, providing objective information regarding the degree of structural damage. A blurred corticomedullary boundary is a pathognomonic characteristic that can be identified by the system with high sensitivity. Measurable shrinkage of the kidney is also an important parameter in the classification algorithm.

For end-stage disease, the system is capable of recognising extremely high and heterogeneous echogenicity, reflecting extensive fibrosis and calcification occurring in end-stage kidneys. Severe cortical atrophy can be detected through complex textural analysis, indicating irreversible damage to the renal structure. Complete loss of corticomedullary differentiation is a hallmark that can be identified with high accuracy. Severe shrinkage of the kidneys provides final confirmation for stage 5 classification.

Real-time Performance and System Output

The system demonstrates excellent real-time performance with the ability to process each image in less than 5 seconds on standard hardware. This processing speed allows seamless integration into clinical workflows without causing bottlenecks in the diagnostic process. High computational efficiency is achieved through CNN architecture optimisation and the implementation of model compression techniques that do not compromise accuracy.

The system output is designed to provide comprehensive information that can be directly utilised by medical personnel. The main prediction label is clearly displayed, accompanied by a confidence level that allows for evaluation of the reliability of the results. The confidence score for each class provides a probability distribution that can assist in differential diagnosis, especially in borderline cases that require more in-depth analysis.

The focus area visualisation feature (attention mapping) shows the regions in the image that most influence the model's decision, providing interpretability that is important for clinical acceptance. This allows medical personnel to understand the basis of the system's decisions and perform visual validation of the analysed areas. Follow-up recommendations based on the classification results provide clinical guidance that can assist in patient management and further therapy planning.

Accuracy of CNN Models in Detecting CKD Characteristics

Model accuracy was evaluated using a comprehensive and objective approach to ensure the validity of the results obtained. The test dataset used was an independent dataset that the model had never seen during the training process, ensuring an unbiased evaluation of the model's generalisation ability in real clinical conditions.

The evaluation strategy applied includes three main approaches. Hold-out validation allocates 20% of the total dataset specifically for testing, providing an objective overview of the model's performance on completely new data. The implementation of 5-fold cross-validation is carried out as additional validation to ensure consistency of results and reduce bias that may arise from specific data divisions. External validation uses external datasets from different sources to test the robustness of the model in real-world conditions with different imaging devices and protocols.

Based on a comprehensive evaluation, the model demonstrated an overall accuracy of 97%. This achievement is the result of optimising the model architecture and a systematic training process, demonstrating excellent pattern recognition capabilities in identifying pathological characteristics in renal ultrasound images.

Detailed analysis shows a training accuracy of 98.5%, indicating that the model can learn complex patterns in the data very well without experiencing underfitting. Validation accuracy reached 97.2%, demonstrating consistent performance on data not seen during training and confirming the absence of significant overfitting. Test accuracy of 97.0% proves excellent generalisation ability, with minimal decline from validation accuracy indicating model stability.

The learning curve analysis shows a stable and consistent pattern of improvement during the training process. The model reached optimal convergence at epoch 35, with progress that can be divided into several distinct learning phases.

a. Epoch 1-10: Rapid improvement in accuracy from 65% to 85%

At this stage, there was a rapid increase in accuracy, from around 65% to 85%. This shows that the model began to effectively recognise basic patterns in kidney images. This is the stage at which the neural network learns the most common features.

b. Epoch 11-25: Stabilisation and fine-tuning reached 94%

After the initial phase, the model underwent a fine-tuning process in which accuracy improvements became more gradual. Accuracy reached 94% and the model began to learn to distinguish more specific features between CKD categories. This phase signalled that the model had begun to stabilise in recognising the characteristics of each class.

c. Epoch 26-35: Final optimisation reached 97%

The model achieves optimal performance with an accuracy of 97%. Convergence is a condition in which the loss function and accuracy no longer experience significant changes between epochs, indicating that the model has found the best weight configuration for the given dataset.

d. Epoch 36-50: Maintenance at an optimal level without overfitting.

At this stage, the model's performance remains stable without any decrease in accuracy or increase in loss on the validation data. This indicates that there is no overfitting, which is a condition where the model memorises the training data too well and fails to recognise new patterns. The model successfully maintains good generalisation.

The model demonstrated excellent performance improvement with accuracy gradually increasing from 65% to 97% at epoch 35. The training process successfully maintained optimal performance until epoch 50 without overfitting, demonstrating good model stability.

To ensure the robustness of the model in diverse clinical applications, testing was conducted on external datasets obtained from different medical institutions with varying device characteristics and patient populations. The results showed consistent performance with 95.8% accuracy, demonstrating good generalisation capabilities despite variations in image quality, ultrasound settings, and patient demographics.

The 1.2% decrease in accuracy during external validation is still within an acceptable range for clinical applications, given the inherent variability in ultrasound images from different sources. These results confirm that the model has sufficient robustness to be implemented in various clinical settings with consistent and reliable performance.

Specificity and Sensitivity Levels of CNN Models in Detecting CKD

Based on the results of training and testing the CNN model on kidney ultrasound images, comprehensive evaluation values were obtained that demonstrated high model performance. The model's accuracy reached 94.69%, which means that the majority of images were classified correctly. A sensitivity of 91.67% indicates that most CKD cases were successfully detected by the system. A specificity of 96% indicates the model's ability to avoid false detections in normal cases. Precision and recall reached 91.67% and 90.91%, respectively, indicating a good balance between prediction accuracy and the ability to detect CKD cases.

The CNN model was trained with a dataset classified based on estimated glomerular filtration rate (eGFR) values, namely normal ($\geq$90 ml/min/1.73 m^2), CKD stage 4 (15–29 ml/min/1.73 m^2), and CKD stage 5 (<15 ml/min/1.73 m^2). The evaluation results showed that the model was able to distinguish between normal kidney conditions and stage 4 and stage 5 CKD with a high and stable level of accuracy.

This classification is based on eGFR values obtained from patient laboratory data, so that each label in the CNN dataset has a valid eGFR value as a standard clinical parameter. The high accuracy of CNN in distinguishing CKD categories based on these eGFR values shows that the model is not only capable of recognising morphological changes in the kidneys on ultrasound, but is also consistent with clinical stratification determined through eGFR.

Although analysis per diagnostic class can provide more detailed insights, this report presents an overall metric evaluation because specific data per class is not explicitly printed from the training results. A more in-depth evaluation can be carried out in future developments by including confusion matrix analysis per normal category, CKD stage 4, and CKD stage 5.

5. Discussion

Capability of a CNN Model in Classifying CKD from Ultrasound Images

The implementation of the Convolutional Neural Network (CNN) model in this study demonstrates significant achievements in the domain of medical image classification, particularly in identifying and categorising Chronic Kidney Disease (CKD) conditions through ultrasound image analysis. The developed model is capable of performing diagnostic differentiation with a high degree of precision across three critical categories: normal kidney condition, stage 4 CKD, and stage 5 CKD, which represent the spectrum of chronic kidney disease progression from a healthy condition to an advanced stage requiring intensive medical intervention.

The fundamental advantage of the CNN approach lies in its ability to automatically perform hierarchical feature extraction, starting from simple edge and texture detection in the early convolution layers, to the recognition of complex morphological patterns in the deeper layers. This process occurs without the need for conventional manual feature engineering, thereby eliminating the subjective bias that often arises in traditional radiological interpretation. This is in line with the modern deep learning paradigm that emphasises end-to-end data representation learning, as explained by Gunawan and Setiawan, who state that CNNs are well suited for medical image data classification due to their ability to perform automatic spatial feature extraction (5).

The implemented CNN architecture demonstrates remarkable adaptability to the inherent variability in ultrasound images, including differences in resolution, contrast, orientation, and noise characteristics generated by ultrasound devices from various manufacturers and generations of technology. This generalisation ability is a crucial indicator for broad clinical implementation potential, given the heterogeneity of medical equipment used in various healthcare facilities.

From a computational learning theory perspective, the model's learning curve shows stable and progressive convergence, starting from a baseline accuracy of 65% and reaching an optimal plateau of 97% at the 35th epoch. Performance stability up to the 50th epoch without any indication of overfitting demonstrates an optimal balance between bias and variance in the model, which is an indicator of the robustness of the learning algorithm. This phenomenon confirms that the model is not only capable of memorising training data, but also truly acquires a deep understanding of visual patterns relevant to CKD diagnosis.

Accuracy of CNN Models in Detecting CKD Characteristics

An overall accuracy of 97% with a training accuracy of 98.5% and a test accuracy of 97.0% demonstrates highly competitive performance, even when compared to conventional diagnostic standards involving manual interpretation by experienced radiologists. These results are in line with the research by Takahashi et al. (2019), which reported an accuracy of 85.6% for CKD classification using ultrasound, and the research by Alharbi et al. (2023), which achieved an accuracy of 97.5% in kidney disease classification using transfer learning. The minimal margin of difference between training and test accuracy (1.5%) indicates the absence of significant overfitting, while also validating the model's generalisation ability towards unseen data.

The CNN model demonstrates superior ability in identifying and analysing key pathological characteristics that are diagnostic markers of CKD, including renal atrophy, cortical thinning, increased parenchymal echo-genicity reflecting tissue fibrosis, and distortion of the corticomedullary junction architecture (23). This multi-parameter detection capability replicates the holistic diagnostic approach used by expert clinicians in daily practice.

A comprehensive evaluation methodology combining hold-out validation, 5-fold cross-validation, and external validation provides a multidimensional perspective on the model's reliability. Hold-out validation provides performance estimates on a separate representative dataset, while k-fold cross-validation optimises training data utilisation and provides estimates of variance in model performance. External validation, which is the gold standard in medical AI validation, tests the robustness of the model against domain shift and population diversity inherent in real-world clinical applications.

External validation results with an accuracy of 95.8% (a decrease of only 1.2% from the test accuracy) demonstrate excellent transferability, indicating that the model has learned generic feature representations and is not tied to the specific characteristics of the training dataset. This is crucial in the context of clinical implementation, where the model will be confronted with patient demographic variations, image acquisition protocols, and equipment characteristics that differ from the development environment.

Specificity and Sensitivity Levels of CNN Models in Detecting CKD

Achieving sensitivity and specificity in initial external evaluations is an important milestone in the field of medical image analysis, given the inherent complexity of interpreting ultrasound images and the high inter-observer variability in conventional radiology practice. A sensitivity of 91.67% indicates that the model is capable of identifying most cases of CKD without producing false negatives, which in a clinical context means a reduced risk of missed diagnoses that could have fatal consequences for patient prognosis. A specificity of 96% demonstrates the model's ability to avoid false positives, which in clinical practice has implications for the prevention of overdiagnosis and unnecessary medical interventions. This balance between sensitivity and specificity is the ideal state in diagnostic testing, given the trade-off that generally occurs between these two parameters in conventional classification systems.

Analysis of the model's performance shows a precision value of 91.67% and a recall of 90.91%, while the F1-score value is in the range of nearly 91%, indicating good and balanced classification performance. The average harmonic between precision and recall expressed in the F1-score confirms that the model is not biased towards certain categories and is able to maintain stable performance.

This success can be attributed to several synergistic technical and methodological factors. First, the CNN architecture was designed with the appropriate depth and complexity for the classification task, avoiding underfitting and overfitting, as suggested in a comprehensive review by Zhang et al. (2024) on CNNs in medical image classification. Second, the careful and clinically relevant implementation of data augmentation techniques, including geometric rotation to reflect natural positioning variations in ultrasound acquisition, spatial translation to accommodate framing variations, and horizontal flipping consistent with the anatomical symmetry of bilateral kidneys, as applied in the methodology of Liu et al. (2022).

Third, a comprehensive regularisation strategy, including dropout layers to prevent co-adaptation between neurons, batch normalisation to stabilise training dynamics, and weight decay to control model complexity (Shorten & Khoshgoftaar, 2019). Fourth, systematic hyperparameter optimisation through grid search or Bayesian optimisation to achieve optimal configurations in learning rate, batch size, and architectural parameters, in accordance with best practices outlined in the transfer learning survey for medical image classification by Alzubaidi et al. (2022).

From a clinical significance perspective, high sensitivity and specificity have major implications for diagnostic workflows and patient outcomes. High sensitivity enables early detection and timely intervention, which are crucial factors in CKD management given the progressive nature of the disease, as emphasised in the National Kidney Foundation guidelines (2024). High specificity reduces the psychological burden on patients due to false positive diagnoses, while optimising resource allocation in the healthcare system by reducing unnecessary follow-up procedures.

Furthermore, the consistency of the model's performance across various validation methods indicates its readiness for clinical implementation and potential integration into radiology information systems (RIS) and picture archiving and communication systems (PACS). This opens up prospects for automated screening programmes, especially in areas with limited access to nephrologists and experienced radiologists, in line with the WHO's vision for universal health coverage and equitable access to healthcare services (Salehin et al., 2023).

In addition to its advantages in detection accuracy, this CNN-based CKD detection system has several technical limitations that need to be considered. First, the bit depth of the ultrasound images used is still within the standard 8-bit grayscale range, which can limit the detail of fine texture information in kidney structures, especially in cases with low contrast. The use of a higher bit depth (e.g., 12-bit or 16-bit) has the potential to improve classification accuracy, but requires a more complex storage and processing system. The next drawback is that full implementation of the system requires a stable internet connection, especially if the application is integrated with a web-based system or cloud server for computing and storing patient data. This condition is particularly challenging in healthcare facilities with limited network infrastructure, so the development of an offline version of the system should be considered to enable wider use.

6. Conclusions

Based on the results of research and analysis conducted on the Convolutional Neural Network (CNN) model in detecting Chronic Kidney Disease (CKD) through kidney ultrasound images, several important conclusions can be drawn. The model demonstrated the ability to accurately distinguish between normal kidneys, stage 4 CKD, and stage 5 CKD. This performance was achieved through the automatic learning of distinctive visual features from ultrasound images, eliminating the need for manual feature engineering. Furthermore, the model achieved an overall accuracy of 94.69%, with a precision of 91.67% and a recall of 90.91%. These metrics reflect the model's effective and balanced capability in detecting renal pathological patterns. In addition, the model reached a specificity of 96% and a sensitivity of 91.67%, indicating its ability to minimize both false positives and false negatives. Such a combination of high specificity and sensitivity is particularly crucial in clinical practice, as it helps to prevent misdiagnosis and ensures accurate early detection of CKD.

References

Aziza, B. L. (2017). *Pemeriksaan struktur dan fungsi ginjal* (pp. 1–51).

Christy, D. (2015). Ultrasonographic assessment of renal parenchymal thickness and kidney length in chronic kidney disease patients. *Journal of Nephrology and Urology, 22*(4), 198–207.

Christy, J., Martadiani, E., & Sitanggang, F. P. (2015). Gambaran ultrasonografi ginjal pada penyakit ginjal kronis berdasarkan stadium di RSUP Sanglah Denpasar. *Jurnal Medika Udayana, 9*(7), 36–40.

Davidson. (2014). *Principles and practice of medicine* (22nd ed.). Elsevier.

Dwi, V., Nursanto, D., Risanti, E. D., Dewi, L. M., Listiana, K., & M. D. (2019). The relationship of hypertension and age against the chronic kidney failure in the Hospital of Dr. Harjono S. Ponorogo. Surakarta.

E. S. (2006). *Nefrologi klinik* (3rd ed.). Bandung.

Gani, N. S. M., Ali, R. H., & Paat, B. (2017). Gambaran ultrasonografi ginjal pada penderita gagal ginjal kronik di Bagian Radiologi FK Unsrat/SMF Radiologi RSUP Prof. Dr. R. D. Kandou Manado periode 1 April–30 September 2015. *e-CliniC, 5*(2). https://doi.org/10.35790/ecl.5.2.2017.17419

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Gunawan, D., & Setiawan, H. (2022). Convolutional neural network dalam citra medis. *Konstelasi: Konvergensi Teknologi dan Sistem Informasi, 2*(2), 376–390. https://doi.org/10.24002/konstelasi.v2i2.5367

Guyton, A. C., & Hall, J. E. (2014). *Buku ajar fisiologi kedokteran* (12th ed.). Penerbit Buku Kedokteran EGC.

Hansen, K. L., Nielsen, M. B., & Ewertsen, C. (2016). Ultrasonography of the kidney: A pictorial review. *Diagnostics, 6*(1), 1–16. https://doi.org/10.3390/diagnostics6010002

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539

Limono, A. B., Setyawan, Y., Tabita, H., & Silitonga, H. (2024). Characteristics and quality of life of chronic kidney disease patients undergoing hemodialysis in Surabaya.

Majdawati, A. (2009). Hubungan gambaran ultrasonografi ginjal dengan laju filtrasi glomerulus (GFR) pada penderita gangguan ginjal: Correlation between renal ultrasound examination and glomerular filtration rate in renal disease patient. *Jurnal Kedokteran Yarsi, 17*(1), 74–81.

Nur Qadar, I. L., Aprianto, N. H., Supriyanto, P., & Aris Diartama, A. A. (2023). Teknik pemeriksaan ultrasonografi panggul dengan klinis kista ovarium di Rumah Sakit Umum Daerah Cengkareng. *Jurnal Ilmu Kedokteran dan Kesehatan, 10*(11), 3141–3147. https://doi.org/10.33024/jikk.v10i11.12722

Rahmayati, E., Sari, G., Apriantoro, N. H., Prayogi, U. D., Irwan, D., Restiyanti, Y., et al. (2021). Gambaran morfologi USG ginjal dengan kreatinin tinggi pada kasus gagal ginjal kronik. *Kocenin Series Conference, 1*(1), 1–7.

Rawat, W., & Wang, Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review. *Neural Computation, 29*(9), 2352–2449. https://doi.org/10.1162/neco_a_00990

Rori, B. A. N., Mamesah, Y. P. M., & Timban, J. F. J. (2024). Gambaran ultrasonografi ginjal pada penderita penyakit ginjal kronis dengan hipertensi di RSUP Prof. Dr. R. D. Kandou periode Juli 2022–Juli 2023. *e-CliniC, 12*(3), 265–270. https://doi.org/10.35790/ecl.v12i3.53469

Sudharson, S., & Kokil, P. (2020). An ensemble of deep neural networks for kidney ultrasound image classification. *Computer Methods and Programs in Biomedicine, 197*, 105709. https://doi.org/10.1016/j.cmpb.2020.105709

Sugiyono. (2022). *Metode penelitian: Kuantitatif, kualitatif dan R&D* (2nd ed.). Alfabeta.

Suwitra, K. (2006). Penyakit ginjal kronik. Dalam Sudoyo, dkk. *Buku ajar ilmu penyakit dalam* (pp. 1–30). Pusat Penerbitan Departemen Penyakit Dalam FKUI.

Tobias, L., Ducournau, A., Rousseau, F., Mercier, G., & Fablet, R. (2016). Convolutional study: Neural network for object recognition on mobile devices. In *International Conference on Pattern Recognition (ICPR)* (pp. 1–6). IEEE. https://doi.org/10.1109/ICPR.2016.7900181

Ummul Aiman, P., Hasda, S., Masita, M., & Sari, M. E. (2022). *Metodologi penelitian kuantitatif*. Yayasan Penerbit Muhammad Zaini.