Indian Journal of Engineering & Materials Sciences
Vol. 33, June 2026, pp. 358-376
DOI: 10.56042/ijems.v33i03.22363

Adaptive CSWin SqueezeNet optimized improved Kepler YOLOv8n with unified framework for ultrasound kidney image retrieval

Sivasakthi Thanigainathan* & Palani Uthirapathy
IFET College of Engineering, Villupuram, Tamil Nadu, 605108, India

Received: 04 August 2025; accepted: 11 April 2026

Kidney stone (KS) detection using ultrasound (US) imaging has become a vital diagnostic tool in urology due to its non-invasive nature and accessibility. However, prevailing models have lacked anatomical context, making KS hard to distinguish from similar conditions. To address these limitations, this research has proposed a novel Adaptive CSWin SqueezeNet-based Improved Kepler YOLOv8n Framework with Unified Image Retrieval (AS-KY-UIR). First, to overcome preconceived notional errors caused by US limitations during augmentation, Artificial Protozoa-optimized Generative Adversarial Network (AP-GAN) has been introduced to dynamically map deviation vectors to improve image quality. For segmentation ambiguity with conditions like Nephrocalcinosis, a hybrid Adaptive Filter-Panoptic CSWin SqueezeNet (AF-PT-SN) model has integrated an adaptive self-guided Filter and panoptic CSWin transformer for enabling effective separation of KS through horizontal-vertical attention. Next, to handle the complexity in KS classification caused by variability in organic matrix compositions, the proposed Improved Kepler optimized Prewitt Relief YOLOv8n Unified Network has used wavelet and relief features to capture gradient-shadow cues, while the Improved Strategy-based Kepler Optimization Algorithm (ISKOA) has refined feature boundaries for precise classification. Finally, to solve annotator bias during image retrieval, a Unified Hierarchical Image Retriever (HIRT) has computed class-wise pixel distances using the Manhattan distance metric for similarity matching. Overall, the AS-KY-UIR framework has enhanced KS detection by improving augmentation with high classification accuracy for supporting timely recovery.

Keywords: Data augmentation, Deep learning, Image retrieval, Kidney stone, Ultrasonic images

1 Introduction

The Kidney Stone Disease, termed as Nephrolithiasis, has rising impacts in recent times and necessitates extensive analysis and multimodal management. Still, the pathophysiology of KS formation is unclear, and the initial step involved is the supersaturation of urine with stone constituents. For identifying the KS, many techniques, namely X-rays, CT scans, and Ultrasound Scans (US), are employed. The US initiates the start of investigation in asymptomatic patients and detects the hydronephrosis and calculi in the kidney. Apart from conventional techniques, many investigations have been carried out using Machine Learning (ML) and Deep Learning (DL) models, which are trained with the images of KS involving pure crystalline composition and KS to attain better performance in medical image recognition. To improve and enrich the training data, augmentation techniques are commonly used to increase the number of image samples in the dataset. In general, augmentation is carried out using horizontal flip, vertical flip, a combination of horizontal and vertical flip, also by adding salt and pepper noises in the dataset images, but these approaches may introduce overfitting and mislead the model, hence improved techniques are preferred[1-5]. Then, for perceiving the Kidney images clearly, pre-processing is carried out, for which the use of a Gabor filter reduces the speckled noises in the US images. By analysing the specific frequency contents of localized regions in images, the filter performs the textural analysis, but faces limitations with edge detection[6].

Further, for achieving effective medical image segmentation, many different algorithms were employed, of which the commonly used one is thresholding. Based on the pixel intensity, the threshold value is chosen and every single pixel value is analogized with the threshold value. The pixel intensity below the threshold value is set to zero, and small changes in the pixel intensity affect the segmentation process[7]. Besides, the U-Net model is used, as it possesses a unique structure with an encoder and a decoder. The defined technique

*Corresponding author (E-mail: sakthi15ifet@gmail.com)

concatenates the features and creates a thick feature, thus preventing information loss, but faces generalization complexities[8]. While in[9], a Convolutional Neural Network (CNN) model is employed for segmenting the images that include colour, shape, and texture, also the ailment is determined, which is carried out by splitting the image into many grids. Every single grid predicts the bounding boxes and identifies the details of renal problems within the grid, but encounters interpretability challenges. Meanwhile, in[10], a ResNEt-18 CNN is employed for general-purpose image segmentation, and it concatenates the demographic and graphical data to attain better accuracy. Nevertheless, the mixed component stones are excluded from the external validations and necessitate quantitative analysis. Further, for extracting the significant features, the grey level co-occurrence matrix (GLCM) statistical tool is employed, also Fuzzy C-means (FCM) and K-means clustering (KMC) are employed, but face complexities in the loss of key characteristics. Moreover, the use of the Restricted Boltzmann Machine (RBM) technique rebuilds the data and extracts the data, but fails to handle gradient blur, hence, novel techniques have to be used[11]. Besides, principal component analysis (PCA) is employed in[12] and it synthesises a series of quantitative features from every image to produce wavelet-based multi-scale features. Then, dragonfly optimization algorithm (DOA) is employed for exact selection of features, for which the attributes and the intellectual ability of the dragonflies are preferred. However, the defined technique faces issues with outliers and local optima[12]. For attaining effective classification of KS, a pre-trained VGG16 model is employed, and it uses 5-fold cross-validation for the analysis, however, the small stones in the lower uterus are challenging and show the lowest recall results[13].

A transformer-based hierarchical learning network architecture is employed for capturing the global and local features, which also involves a multi-head attention layer for attaining the long-time dependencies[14]. Similarly, for classifying the regions, a random forest classifier (RGC) is employed, and it uses a tree depth of 9 splits before the forced leaf. When this classifier is combined with U-Net gives out better classification, however faces issues with respect to rejection of incomplete insertion classes[15]. Further, for effective image retrieval, an unsupervised content-based medical image (UCMI), which uses spatial matching of visual words, is presented. The UCMI significantly computes the spatial similarity using the skip similarity index, but the location features match only if the difference between the locations is less than 20%[16]. Meanwhile, an image retrieval method based on image feature fusion is carried out, and it employs object-centric and place-centric deep networks for extracting the necessary features and gives out better dimensional feature reduction. However, necessitates better enhancement in a defined model in case of larger datasets[17].

Radu Alexa *et al.*[18] and Balangeshwara *et al.*[26] employed CNN-based architectures to automate the ultrasound KS detection, improving detection performance and diagnostic precision. But the method is highly sensitive to variations in image quality, patient variations, and low-resolution images. Similarly, Mohan Bhandari *et al.*[19] developed a lightweight customized CNN with additive Shapley value-based explanation for KS detection, providing a better interpretability and target localization, while necessitating improved augmentation. A microbiome-based detection approach proposed by Ava Mqusavi *et al.*[20], which utilized the microbial DNA sequencing and statistical analysis to extract disease-related insights. However, the method lacks sufficient patient data and gender differences. Further, Dina Salif *et al.*[21] and Deema Mohammed Alsekait *et al.*[22] introduced ensemble learning frameworks with CNN, LSTM, and feature selection which enables higher classification performance. However, these are limited by class imbalance, high-dimensional features, and a lack of correlation analysis.

Similarly, for KS localization, Tirupal *et al.*[23] developed a deep semantic segmentation architecture based on image characteristics, but low contrast and noise in ultrasound images worsen the segmentation accuracy. Meanwhile, Asaye *et al.*[24] built a ML framework for nephrolithiasis deciohm from ultrasound images using GLCM features with SVC, DT, KNN, and RF classifiers, but the reliance on handcrafted features suppressed by image variation, annotation bias, and poor dataset generalization. While, Jadhav *et al.*[25] introduced a hybrid CNN-LSTM model for KS detection and classification model, which promotes diagnostic precision as well as clinical decision-making, but they face issues with overfitting, patient-specific variations, and low-

resolution images. Finally, Mahalakshmi *et al.*[27] presented an ensemble transfer learning approach via DCNN models. The system's reliance on pre-trained models hampered its capacity to adapt to domain-specific ultrasound picture data and the computational complexity. However, these methods are limited by ultrasound variability and computational complexity, which results in limited compatibility for real-time clinical implementation.

From the discussion, it is concluded that KS detection using ultrasound imaging has been improved through conventional approaches, but several challenges remain. Existing methods, involving pre-processing, segmentation, feature extraction, classification, and image retrieval, encounter limitations such as speckle noise, edge detection, pixel intensity variations, feature loss, gradient blur, local optima, low recall, spatial matching, and scalability issues. Moreover, ML and DL-based approaches are affected by ultrasound variability, image quality variations, patient-specific differences, annotation bias, class imbalance, overfitting, generalization and interpretability issues, which limit practical clinical deployment. From the analysis of the identified challenges, the motivation for the research work for the ultrasound kidney image retrieval is established, which is explained below.

1.1 Research motivation

The augmentation process involved in KS detection highly depends on preserving anatomical context. However, ultrasound scan (US) struggle to visualize deep structures, due to the variations in fat content and tissue density resulting in attenuation-based Comet tail artifacts (bright line of echos). Hence, the deviations from the normal image increases and appear as severe pathological conditions, causing domain shifts. This results in preconceived notions, which cascades through all the levels of analysis and leading to missed symptoms or misinterpretation of diagnostic results.

Further for segmentation, most of the algorithms treat images as collections of pixels and ignore the broader context of the image, reducing the segmentation accuracy. Even though recent techniques consider the contextual information during segmentation, KS identification remains challenging due to the presence of Nephrocalcinosis. These deposits can mimic KS, especially in the medulla region, creating ambiguous feature that resuts in closure loss and uncertainty in KS extent estimation. In KS classification, the heterogeneous nature of organic matrix (stone compositions), creates challenges in identifying CaOx (Calcium Oxalate) structures. The organic matrix includes Promoters and Inhibitors, where the imbalance constitutes stone formation with different sizes ranging (<5 mm, 5-10 mm, 10-20 mm, and >20 mm), which is still a challenging issue for the conventional techniques to classify. When the promoters outweigh the inhibitors, the interwoven connection of factors in the organic matrix differs and it distorts the edges of the posterior acoustic shadow of the stone (which computes the size of stone in the US), leading to overestimation of stone sizes. Hence, determining the exact size, leads to unnecessary interventions about the severity (i.e.) mild, moderate and severe and also extends the recovery time.

Besides, during image retrieval, due to ambiguity, the fine-grained details overlook important features and leads to Annotator Bias, in which the training data reflects, reinforces and perpetuates existing biases. This influences the data labelling process, leading to defective annotations. To overcome these issues in KS detection and image retrieval, this research develops a novel DL-based optimized architecture for effective retrieval and classification of kidney ultrasound images. Additionally, the primary contributions of this paper are as follows,

a To propose the AS-KY-UIR framework integrating AP-GAN, AF-PT-SN, and Kepler optimized YOLOv8n for enhanced KS segmentation and retrieval using ultrasound images.

b To design an Adaptive Filter-Panoptic CSWin transformer-based Squeeze Net to effectively isolate and segment ambiguous KS features from nephrocalcinosis in noisy ultrasound data.

c To introduce an Improved Kepler optimized Prewitt Relief YOLOv8n Unified Network that combines multi-scale edge detection, relevance-based feature selection, and optimized classification for precise KS localization, grading, and unified image retrieval.

d To implement a Unified Hierarchical Image Retriever combining generative and retrieval-based approaches to reduce annotator bias and improve fine-grained image similarity analysis.

2 Materials and Methods

As the conventional KS identification techniques faces issues in terms of anatomical context in

augmentation and stone composition in classification, a novel "Adaptive CSWin Squeeze Net-based Improved Kepler YOLOv8n Framework with Unified Image Retrieval (AS-KY-UIR)" model is proposed in this research.

Initially, for diversifying the KS dataset and to enhance the accuracy of the optimized DL techniques used, image augmentation is carried out. As the ability of US lags in visualizing the deep structures, a novel artificial protozoa optimized generative adversarial network (AP-GAN) is employed. The GAN used involves a generator and a discriminator framework for learning the data and to synthesise new realistic images from it. The generator produces images alike the real images, by mimicking the patterns and structures of the trained data, which reduces the come tail artifacts effectively. This effectively fools the discriminator to differentiate the original images and hence for improving its performance, the discriminator adapts a bio-inspired Artificial Protozoa optimization algorithm. The algorithm uses a mapping vector to identify the dimensional crossover of the candidate solutions, which detects the level of deviations from the original image. Based, on this the discriminator sends the feedback signal to the generator, which reduces the bias in preconceived notion, thereby aids in enhancing the quality of data that has to be used during the classification.

In order to segment the KS features effectively, the input image taken is processed using a novel adaptive filter-panoptic CSWin transformer-based squeezenet (AF-PT-SN) Framework. The Squeezenet Framework employed for Segmentation involves two stages namely pre-processing and segmentation at its hidden layers. Initially, the input image is pre-processed using a novel Contrast Adaptive Self-Guided Pixel Wise Filter. The Self-Guided Pixel Wise Filter employed uses a guidance image and with its reference, the redundant noises in the image are removed and conserve the textural details. Further, for improving the contrast, a Contrast Adaptive Histogram Equalization method is employed, in which the intensity values of the images are redistributed by dividing the image into smaller regions and using contrast limiting mechanism, the contrast is enhanced with effective prevention of extreme changes. Then, for segmenting the features necessary for classification, Panoptic CSWin transformer based Segmentation (P-CSWinT) is employed. The CSWin uses horizontal and vertical self-attention mechanisms for identifying the calcium deposits and the content-aware reassembly operator, which is guided by predicted kernels, assists in identifying the hyperechoic areas due to Primary Hyperoxaluria. Further, segmentation is carried out using Panoptic Segmentation technique that combines the semantic and instance segmentation. The defined technique segments the features using two labels, in which the semantic label isolates the features and the instance label distinguishes the different instances amid the same classes, which brings effective segmentation of the KS from the Nephrocalcinosis. With this, the ambiguous feature formed by Nephrocalcinosis that creates closure loss is excluded and the necessary KS information is captured and segmented.

For effective classification and retrieval of kidney images, a novel Improved Kepler optimized Prewitt Relief YOLOv8n Unified Network is proposed, which includes an input layer, three hidden layers and an output layer, where feature extraction, selection and classification stages, also image retrieval is carried out. At the first hidden layer, feature extraction is carried out, for which, prewitt wavelet transform is employed, in which the wavelet transform decomposes the images to different frequency bands, thereby gives better resolution. Meanwhile, the prewitt operator in it, which is a discrete differentiation operator, extracts the interwoven connection of factors in the organic matrix using two convolutional kernels, which approximates the gradient of image intensity function. Then, at the second hidden layer, using iterative relief feature selector (IRFS), the feature variations in the organic matrix is identified, for which the algorithm uses a ranking method. Based on the sample chosen, the proximity of other classes, also its distance from the original class is noted (i.e. the distance amid the acoustic shadow and the stone is noted). This, thereby aids in identifying the feature dependencies and based on this the feature selection is carried out. Further, to pin-point the exact size of the stone and to classify it, an improved strategy based kepler optimization algorithm (ISKOA) is employed at the third hidden layer and it uses a good point set strategy to ensures that the initial solution (posterior acoustic shadow) is evenly distributed across the domain. Then, for reducing the overestimation of the stones, a dynamic opposition based learning technique is employed and

it selects the higher fitness value using a boundary mechanism (i.e.) the appropriate extent of the shadow. Further, based on novel location update strategy, the severity of the KS (i.e.) mild, moderate and severe conditions are identified.

Further, the classified output from the third hidden layer is fed to the output layer, in which a Unified Hierarchical Image Retriever is employed for eliminating the Annotator bias. In order to identify the fine grained details of the important features, a unified framework that involves a generative Retrieval Method is used and it unifies the generation and retrieval, with the aid of an autonomous decision module, which further reduces Annotator bias. Further, as a response to query, the best matches amid the retrieved and the generated images are selected. Then, for reducing the impacts of data labelling process and to improve the image retrieval with better similarity, a hierarchical image retrieval technique (HIRT) is employed and it combines visual and semantic similarity. The HIRT involves two stages, in which the first stage involves image level feature vectors and is stored as reference, while in the second stage, the class of every single pixel with the query and the candidate image is identified. Then, using Manhattan distance matric, the distance amid the class and the query image is identified, thereby reducing the defective annotations. Hence, with the novel DL based optimized techniques, the KS image retrieval is carried out effectively, limiting the unnecessary interventions and assists in reducing the recovery time.

The AS-KY-UIR methodology for ultrasound kidney image retrieval in Fig. 1 begins with the input ultrasound image (a query image) being passed into the AF-PT-SN framework. Here, preprocessing is carried out using a contrast adaptive self-guided pixel-wise filter to remove noise while preserving essential fine details. Following this, segmentation is performed using a Panoptic CSWin Transformer-based approach, which accurately differentiates between KS and other hyperechoic structures by combining semantic and instance segmentation strategies. The segmented outputs are further processed by the improved Kepler optimized prewitt relief YOLOv8n unified network block, which begins with prewitt wavelet transform for robust feature extraction by capturing critical edge information related to the organic matrix structure of stones. An iterative relief feature selector is then applied to

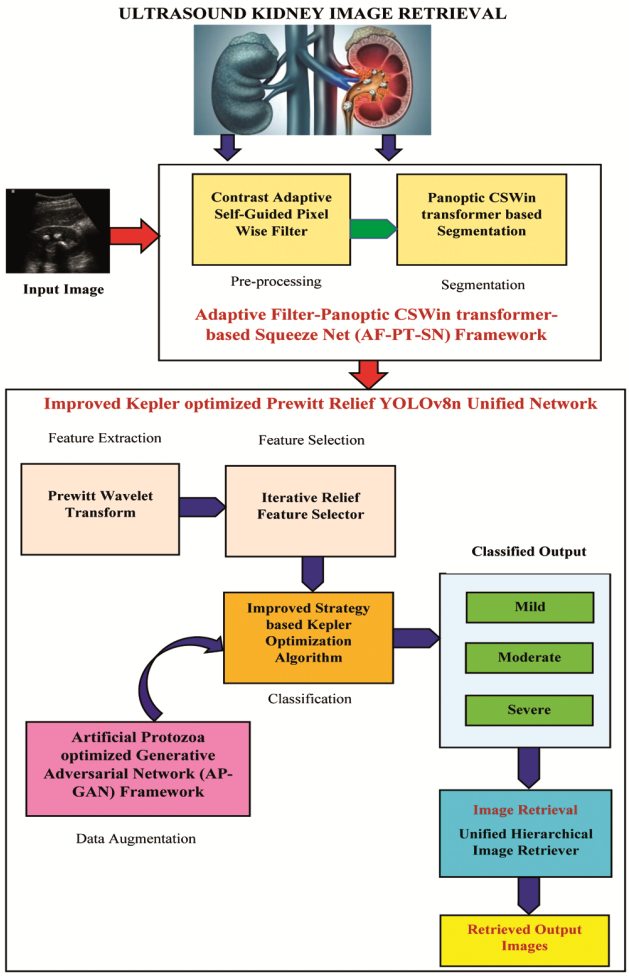


Fig. 1 — Block structure of AS-KY-UIR for ultrasound kidney image retrieval and classification.

prioritize the most discriminative features that separate stones from surrounding tissues. This block contains an enhanced augmented dataset, which is created initially by using an AP-GAN framework to address artifacts and improve image quality for better downstream processing of classification. The selected features of the query image are classified using an improved strategy-based kepler optimization Algorithm, which applies dynamic and opposition-based learning mechanisms to precisely classify KS into three severity levels, namely mild, moderate, and severe. Finally, the classified output image is used in a unified hierarchical image retriever, which systematically retrieves the most semantically similar ultrasound image, thereby ensuring consistent labeling, minimizing annotator bias, and improving the reliability of the retrieval process for diagnostic decision support.

2.1 Artificial protozoa optimized generative adversarial network (AP-GAN)

To strengthen the diversity of the KS dataset and boost the performance of the optimized deep learning models, an effective image augmentation strategy is applied. Since US imaging often struggles to reveal deeper anatomical details, the proposed work introduces an AP-GAN to overcome these limitations, while the working is analyzed in Fig. 2.

Within this framework in Fig. 2, the generator is tasked with creating new synthetic images that replicate the structural and textural details found in the original scans, thus helping to minimize common imaging issues like comet tail artifacts. By carefully learning the intricate patterns of the kidney regions, the generator produces high-quality augmented images, ensuring that even subtle features are preserved while distortions are significantly reduced. This augmented data subsequently strengthens the dataset, enabling the classification models to learn from clearer and more accurate examples.

In parallel, the discriminator in the AP-GAN is strengthened using the Artificial Protozoa Optimization (APO) technique to enhance its decision capabilities, a biologically inspired movement is determined mathematically by using equation (1)[28].

$$Z_i^{new} = Z_i + F_{fac} \cdot [Z_j - Z_i + {}^1/_{mq} \cdot \sum_{n=1}^{mq} Y_x \cdot Zh - - Zh + \odot Rmap \quad \ldots (1)$$

This formulation in Equation (1) governs how each candidate solution "Z_i", represents a possible decision boundary of the discriminator and updates its position in response to environmental feedback. Here, "Z_i^{new}" denotes the newly updated position of the i^{th} candidate and "F_{fac}" is a force factor that modulates the magnitude of movement based on interaction strength. The term "$Z_j - Z_i$" captures the local influence from another randomly selected candidate "Z_j", encouraging diversity. The summation term "$^1/_{mq} \cdot \sum_{n=1}^{mq} Y_x \cdot (Z_{h-} - Z_{h+})$" aggregates the directional influence from multiple high-fitness "$h-$" and low-fitness "$h+$" candidates, with "Y_x" being the protozoa response coefficient controlling the influence of each neighbor. Finally, "$\odot R_{map}$" introduces a random mapping vector R_{map}, which dynamically perturbs the update path, promoting exploration and reducing the risk of local optima. This update mechanism allows the discriminator to effectively track the real image distribution boundaries, setting the stage for improved feedback to the generator and more realistic image synthesis. Further, Equation (2) introduces a localized update strategy to refine the discriminator's search path.

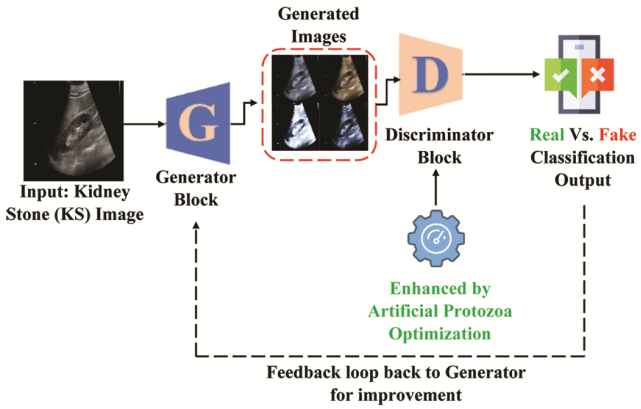


Fig. 2 — Functional Block of AP-GAN network.

$$Z_i^{new} = Z_i + F_{fac} \cdot [Z_{near} - Z_i + {}^1/_{mq} \cdot n{=}1mqYx \cdot Zi - h - Zi + h \odot Rmap \quad \ldots (2)$$

In this, "Z_i^{new}" is the refined candidate position, "Z_i" is the current solution, and "Z_{near}" refers to its nearest neighbor guiding the adjustment. The term "$^1/_{mq} \cdot \sum_{n=1}^{mq} Y_x \cdot (Z_{i-h} - Z_{i+h})$" incorporates influence from surrounding candidates at a distance "h", helping balance exploration and exploitation. "F_{fac}" tunes the update strength, while "R_{map}" adds randomness for adaptability. This localized interaction sharpens the discriminator's focus on subtle distribution gaps, ensuring more accurate differentiation between real and generated samples. Rather than traditional updating methods, the discriminator employs a dynamic mapping vector that analyzes candidate solutions, identifying any deviations from the real image distributions.

Moreover, Equation (3) defines a boundary-aware initialization used to diversify the discriminator's search space.

$$Z_i^{new} = Z_{min} + R_{ran} \odot (Z_{max} - Z_{min}) \quad \ldots (3)$$

Here, "Z_i^{new}" represents the newly initialized solution, bounded within the range "$Z_{max} - Z_{min}$". The term "R_{ran}" is a randomly generated vector controlling the position within this boundary, while "$\odot$" indicates element-wise multiplication. This mechanism ensures that the initial candidates are evenly and randomly distributed across the feasible search domain, helping the discriminator avoid premature convergence and better explore the

distribution of real versus generated images. Additionally, through this mechanism, the discriminator continuously refines its decision boundaries and provides critical feedback to the generator, pushing it to generate even more realistic and less biased outputs, which is derived mathematically by using Equation (4).

$$Z_i^{new} = Z_i \pm R_{ran} \cdot [Z_{min} + R_{ran} \odot (Z_{max} - Zmin \odot Rmap \quad \ldots (4)$$

As a result, the generator becomes increasingly effective at eliminating noise and artifacts like comet tails, ultimately yielding a higher quality training set, which is determined clearly by examining algorithm 1.

Algorithm 1: Artificial Protozoa Optimization

Start

Input:

a. Initial population of candidate discriminator boundaries $\{\boldsymbol{Z_i}\}_{\boldsymbol{i=1}}^{\boldsymbol{N}}$
b. Parameters: population size "$\boldsymbol{N}$", number of elite solutions "$\boldsymbol{mq}$", movement step, force factor "$\boldsymbol{F_{fac}}$", random maps "$\boldsymbol{R_{map}}$", $\boldsymbol{R_{ran}}$, search bounds $\boldsymbol{Z_{min}}$ and $\boldsymbol{Z_{max}}$.

Output:

a. Optimized discriminator boundaries $\boldsymbol{Z_i^{new}}$ for improved decision making.
b. Refined feedback to generator for enhanced augmentation.

Step 1: Initialization

i. For each candidate $\boldsymbol{Z_i}$, initialize using boundary-aware strategy using Equation 3.

Step 2: Global Interaction Update

i. For each candidate $\boldsymbol{Z_i}$, update position using global protozoan motion using Equation 1.

Step 3: Localized Fine-Tuning

i. Apply local refinement using nearest neighbor influence using Equation 2.

Step 4: Boundary-aware Perturbation

i. Further enhance with dynamic boundary perturbation using Equation 4.

Step 5: Evaluate Fitness

i. Evaluate each updated $\boldsymbol{Z_i^{new}}$ based on discriminator's classification accuracy against real vs. generated distributions.

Step 6: Selection and Replacement

i. Retain high-fitness candidates for next generation and replace weaker ones with refined versions.

Step 7: Repeat Until Convergence

i. Repeat Steps 2-6 until stopping criteria met (max iterations or fitness threshold reached).
ii. Return the best $\boldsymbol{Z_i^{new}}$.

End

The algorithm begins by initializing candidate solutions within defined bounds to ensure a diverse search space. Through global interaction by Equation 1, each candidate adjusts its position based on the influence of stronger and weaker peers, promoting exploration. A localized refinement step by Equation 2 then sharpens this search by considering nearby candidates. Further diversity is introduced using boundary-aware initialization by Equation 3 and dynamic perturbation strategies by Equation 4, helping avoid local optima. This continuous refinement enables the discriminator to provide stronger feedback to the generator, leading to the production of high-quality, noise-free images that improve classification reliability in the AS-KY-UIR model. Therefore, the improved augmentation pipeline directly contributes to better classification reliability and accuracy across the subsequent stages of the AS-KY-UIR model.

2.2 Adaptive filter-panoptic CSWin transformer-based squeeze net (AF-PT-SN) framework

Accurate segmentation of KS features is often challenged by the ambiguous patterns formed due to Nephrocalcinosis, leading to closure loss. To overcome this, a novel AF-PT-SN framework is proposed, which is given in Fig. 3 and Fig. 4.

The SqueezeNet framework utilized for segmentation operates in a dual-stage process, encompassing both pre-processing and segmentation within its hidden layers. The integration ensures context-aware delineation of structures, significantly improving segmentation precision, where a detailed working is given in Fig. 4.

In the hidden layers of squeeze net initially, a Contrast Adaptive Self-Guided Pixel-Wise Filter refines the input by removing noise while preserving textures. Further enhancement is achieved using Contrast Adaptive Histogram Equalization to improve local contrast without over-amplifying artifacts. For feature extraction, a P-CSWinT employs directional

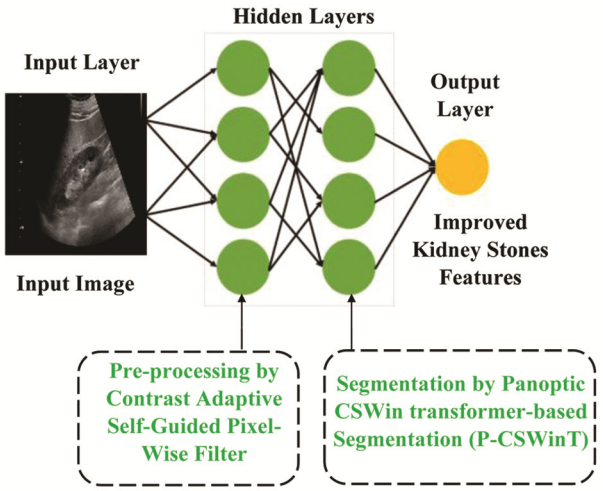


Fig. 3 — Context-aware dual-stage segmentation using squeezenet background.

self-attention to precisely detect calcium deposits and hyperechoic regions. The panoptic segmentation then separates semantic regions and instances, ensuring KS structures are distinctly isolated. Through this, the influence of Nephrocalcinosis is minimized, enabling clearer and more reliable KS segmentation for producing efficient purification of KS images.

2.2.1 Contrast adaptive self-guided pixel-wise filter

The Contrast Adaptive Self-Guided Pixel-Wise Filter is designed to enhance ultrasound images by intelligently removing noise while preserving important textural details. It works by first generating a guidance image that acts as a clean reference to direct the filtering process. Instead of applying a fixed smoothing across the entire image, the filter adapts its behavior pixel-by-pixel based on local contrast variations, ensuring that fine structures are maintained while unwanted distortions are suppressed and determined precisely by using Equation (5)[29].

$$Z[x,y] = \sum_{i\in k}[(xI_i + y - N_i)^2 + X'x^2] \quad \ldots(5)$$

In Equation (5), "$Z[x,y]$" represents the filtered output value at each pixel position "$[x,y]$", capturing the refined intensity after contrast-based adjustment. The summation runs over all neighboring pixels within a local window "k" and helps the filter to assess the local structure around each pixel. "I_i" denotes the intensity of the neighboring pixel, while "N_i" refers to the intensity value from the guidance image, acting as a clean reference for comparison. The term $(xI_i + y - N_i)^2$ measures the squared difference between the combined weighted pixel values and the guidance image, reflecting local noise or distortion levels. The additional component "$X'x^2$" is a contrast-aware regularization term that penalizes excessive changes, with "X'" representing a scaling factor based on local contrast. Together, these terms enable the filter to dynamically adjust smoothing strength based on the texture and contrast of different regions, preserving important features while removing noise effectively. Hence, regions with higher texture of important features are preserved more carefully, while flatter areas prone to noise are more strongly filtered. This pixel-wise self-guidance helps to achieve a balanced enhancement where clarity is improved without blurring the essential KS features. As a result, the processed image provides a stronger foundation for accurate P-CSWinT segmentation in the following stages.

2.2.2 Panoptic CSWin transformer-based segmentation (P-CSWinT)

To accurately extract the necessary filtered features for reliable KS segmentation, the proposed method adopts a P-CSWinT framework. This technique capitalizes on the strengths of the CSWin Transformer, which applies a shifted window-based self-attention mechanism both horizontally and vertically across the image by using Equation (6) and (7)[30].

$$z_n = \begin{cases} H_{atten_n}[Y], n = 1,2,\ldots\ldots, {}^{N}/_{2} \\ V_{atten_n}[Y], n = {}^{N}/_{2} + 1, {}^{N}/_{2} + 2, \ldots\ldots, N \end{cases} \quad \ldots(6)$$

$$CSWin_{atten}[Y] = z_{concat}[z_1, z_2, \ldots\ldots\ldots z_N]w_{matrix} \quad \ldots(7)$$

where, "z_n" in Equation (6) represents the attention output for the n^{th} partition of the feature map "Y", capturing either horizontal or vertical context. For the first half of the partitions "$n = 1,2,\ldots\ldots, {}^{N}/_{2}$", "$H_{atten_n}[Y]$" applies horizontal self-attention across shifted windows, helping detect elongated structures like calcium streaks. For the second half "$n = {}^{N}/_{2} + 1, {}^{N}/_{2} + 2, \ldots\ldots, N$", "$V_{atten_n}[Y]$" applies vertical self-attention to highlight vertically aligned textures like deposits or lesions. Moving to Equation (7), "$CSWin_{atten}[Y]$" denotes the combined CSWin attention output. Here, the outputs from each partition "$[z_1, z_2, \ldots\ldots\ldots z_N]$" are concatenated together, forming "z_{concat}", which is then multiplied with a learnable weight matrix "w_{matrix}" to refine the final attention-enhanced features.

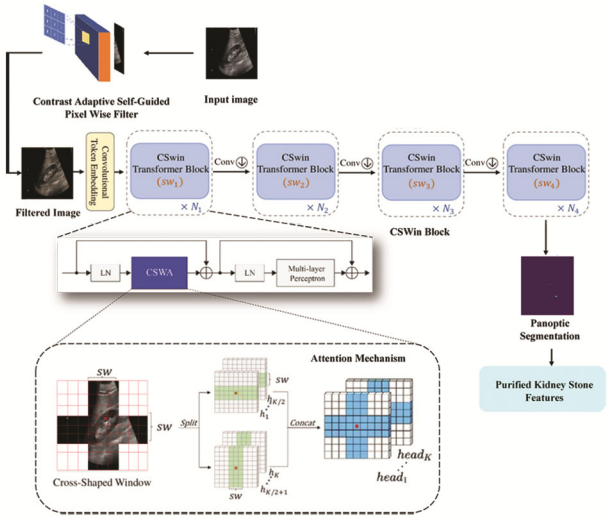


Fig. 4 — Filtered segmentation process of AF-PT-SN in AS-KY-UIR framework.

Such dual-directional attention enables the model to capture intricate contextual information and long-range dependencies, which are vital for identifying calcium deposits hidden within complex anatomical structures. Unlike conventional transformers that often overlook local details, the CSWin model selectively focuses on spatial regions with higher diagnostic significance, ensuring that minor yet critical features like stone formations are not missed. Additionally, the integration of a content-aware reassembly operator, which functions under the guidance of dynamically predicted kernels, significantly enhances the model's ability to reconstruct finer structures, which is given by Equation (8).

$$Z'_{X'} = \emptyset[N(Z_x, K_{size}), C_{X'}] \quad \ldots(8)$$

where in Equation (8), "$Z'_{X'}$" represents the reconstructed feature map after reassembling finer structural details from the input. The function "$\emptyset$" symbolizes the content-aware reassembly process, which intelligently pieces together important features. "$N(Z_x, K_{size})$" denotes the neighboring region extraction around a pixel "x", where "K_{size}" defines the size of the dynamic kernel predicted specifically for each pixel. This kernel adapts to local content, allowing precise focus on fine textures like small calcifications. Meanwhile, "$C_{X'}$" is the context information extracted from the input feature map "X'", guiding how the neighboring regions are combined. This is particularly important for accurately highlighting hyperechoic regions, often indicative of Primary Hyperoxaluria, thus improving the differentiation of pathological features. After this attention-driven extraction, the framework proceeds with a Panoptic Segmentation strategy that fuses semantic and instance segmentation into a unified prediction. The semantic segmentation label identifies and isolates broader regions of interest, ensuring the stones are cleanly separated from normal tissue backgrounds. Meanwhile, the instance segmentation label works to distinguish individual stones, even when multiple instances are present within the same tissue class. This dual-labeling approach is highly effective in mitigating the challenge posed by ambiguous features, especially those arising from Nephrocalcinosis, which often resemble KS in ultrasound imagery. The precise segmentation results from P-CSWinT lay the foundation for the next classification stage, where only the necessary and purified KS features are analyzed for better diagnostic accuracy.

2.3 Improved kepler optimized prewitt relief YOLOv8n unified network

To enhance the classification accuracy and efficient retrieval of kidney images, an advanced Improved Kepler optimized Prewitt Relief YOLOv8n Unified Network is introduced. This framework is structured with an input layer, a sequence of three hidden layers, and a final output layer to streamline the entire process. The hidden layers collaboratively perform feature extraction, critical feature selection, and robust classification of the input data.

The overall architecture of the proposed unified network, as implemented in this study is shown in Fig. 5. The first hidden layer of the unified network initiates the feature extraction process using the Prewitt Wavelet Transform, which fuses the power of wavelet decomposition with the edge-detection capabilities of the Prewitt operator, depicted by Equations (9) and (10)[31].

$$P_{hor} = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} * f(a,b) \quad \ldots(9)$$

$$P_{ver} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} * f(a,b) \quad \ldots(10)$$

In Equations (9) and (10), "P_{hor}" and "P_{ver}" represent the horizontal and vertical edge feature maps extracted from the input kidney ultrasound image, respectively. Here, the matrices shown are the

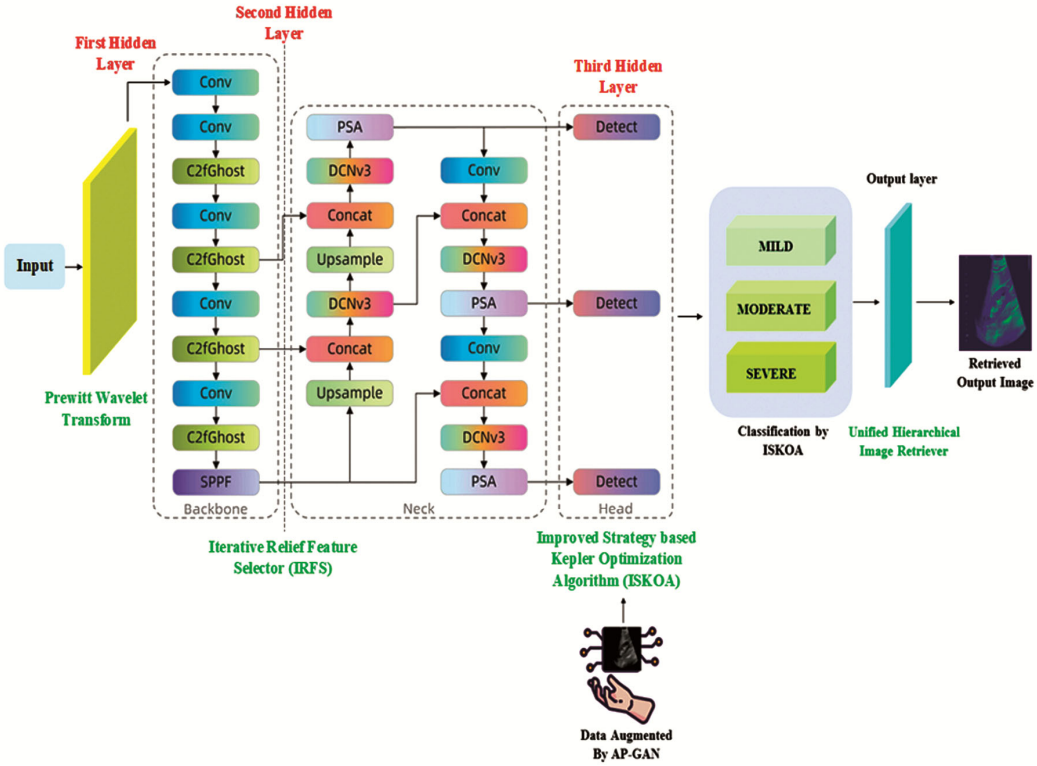


Fig. 5 — Detailed structural view of improved kepler optimized prewitt relief YOLOv8n unified network for classification.

Prewitt operators, where the horizontal filter emphasizes differences along rows (detecting vertical edges), and the vertical filter highlights differences along columns (detecting horizontal edges). The symbol "$*$" indicates the convolution operation applied over the input image "$f(a,b)$", where "a" and "b" denote the pixel coordinates. Further, Wavelet decomposition breaks down the ultrasound image into multiple frequency bands, allowing the model to analyze both fine and coarse textures simultaneously capturing subtle variations often missed by traditional filters. This multi-resolution approach helps preserve essential structural patterns crucial for KS analysis. Embedded within this is the Prewitt operator, a gradient-based technique that uses directional convolutional kernels to detect intensity changes in both horizontal and vertical directions. This dual mechanism not only enhances the clarity of stone boundaries with different size ranges like <5 mm, 5-10 mm, 10-20 mm, and >20 mm, but also extracts the intricate relationships within the organic matrix, including promoters and inhibitors that influence stone formation. These extracted gradients form the foundation for the subsequent feature selection stage, where the model evaluates which attributes are most critical for accurate classification.

In the second hidden layer IRFS is used to refine the features extracted from the previous stage by identifying those most relevant for classification. Unlike conventional selectors that simply reduce dimensionality, IRFS intelligently evaluates the contribution of each feature by analysing how well it distinguishes between classes particularly in the presence of overlapping patterns, such as the acoustic shadow and stone region. For each input sample, IRFS calculates the distances to its nearest neighbours from the same and different classes, enabling the algorithm to assess how strongly a feature contributes to the correct classification. This iterative ranking mechanism allows the model to uncover subtle dependencies in the organic matrix, ensuring that only the most discriminative features are retained. These ranked features, filtered for their class-separating power, are then passed on to the optimization layer, where precise size classification of the KS is further enhanced.

At the third hidden layer, the ISKOA is employed to accurately estimate the size (<5 mm, 5-10 mm, 10-20 mm, and >20 mm) of KS and classify them based on severity with the help of augmented data from AP-GAN. This algorithm tackles one of the major challenges in ultrasound imaging, where the

overestimation caused by distorted acoustic shadows is solved by initiating the optimization process with a good point set strategy, which ensures that the search space is thoroughly and evenly explored using Equation (11) [32].

$$F_g(t) = E_{ecc} \times \mu(t) \times {}^{\bar{M}_n \times \bar{m}_i}\Big/_{\bar{N}_i^2 + \lambda} + R_{ran} \quad \ldots (11)$$

where, "$F_g(t)$" in Equation (11) denotes the global fitness value at iteration "t", which helps guide the search towards the optimal KS size classification. Here, "E_{ecc}" represents the eccentricity energy used to control exploration, "$\mu(t)$" is a dynamic control factor that adjusts over iterations to balance exploration and exploitation, "$\bar{M}_n$" and "$\bar{m}_i$" are the normalized maximum and individual fitness values, and "$\bar{N}_i$" is the normalized distance metric defined in Equation (12).

$$\bar{N}_i = {}^{N_i(t) - min[N(t)]}\Big/_{max[N(t)] - min[N(t)]} \quad \ldots (12)$$

In Equation (12), "$\bar{N}_i$"is calculated by rescaling the distance "$N_i(t)$" between solutions relative to the minimum and maximum distances at the current iteration, ensuring fair comparison. "λ" is a regularization term to prevent division errors and improve stability. Instead of relying on a random or biased starting point, this strategy distributes initial solutions across the entire domain, improving the likelihood of converging on the true size of the stone. To further refine this estimate, dynamic opposition-based learning is integrated, which compares each solution with its opposite value to select the one with higher fitness, thereby effectively correcting size inflation. Additionally, a novel location update mechanism is utilized to dynamically adjust the position and extent of the posterior acoustic shadow using Equation (13), enabling precise categorization of stone severity as mild, moderate, or severe.

$$\vec{Z}_i[t+1] = \vec{Z}_i[t] + \sigma \times \vec{V}_i[t] + \vec{U} \times \left[F_g(t) + Rran \times Zst - Zit\right] \quad \ldots (13)$$

Finally, Equation (13) updates the position of the solution "$\vec{Z}_i[t+1]$", where "σ" controls the influence of velocity "$\vec{V}_i[t]$", "$\vec{U}$" is a uniform random vector that adds stochastic behavior, "R_{ran}" is the random component and "$\vec{Z}_s[t]$" is the best solution found so far. Together, these equations help ISKOA efficiently explore the search space and accurately predict KS sizes despite imaging artifacts as studied using algorithm 2.

Algorithm 2: Improved Strategy based Kepler Optimization

Start

Input:

a. Extracted feature set from previous layers (Prewitt Relief YOLOv8n output).
b. Population size $\boldsymbol{P}$, maximum iterations $\boldsymbol{T}$.

Output:

a. Predicted Kidney Stone Size Category (<5 mm, 5-10 mm, 10-20 mm, >20 mm) as mild, moderate, or severe.

Step 1: Initialization

Initialize a random population of candidate solutions $\vec{\boldsymbol{Z}}_{\boldsymbol{s}}[\boldsymbol{0}]$ representing possible classifications.

Initialize velocities $\vec{\boldsymbol{V}}_{\boldsymbol{i}}[\boldsymbol{0}]$ and control parameters $\boldsymbol{\sigma}$, $\boldsymbol{\lambda}$, and $\boldsymbol{\mu(0)}$.

Compute initial fitness values $\boldsymbol{F_g(0)}$ for each candidate using extracted features.

Step 2: Normalize Distances

For each candidate, compute the normalized distance $\bar{\boldsymbol{N}}_{\boldsymbol{i}}$ using Equation (12).

Step 3: Fitness Evaluation

Evaluate the global fitness $\boldsymbol{F_g(t)}$ for each candidate using Equation (11).

Here, $\boldsymbol{E_{ecc}}$ adjusts exploration, $\bar{\boldsymbol{M}}_{\boldsymbol{n}}$ is the best fitness in the population, and $\boldsymbol{R_{ran}}$ maintains diversity.

Step 4: Update Solutions

Update each candidate's position using Equation (13).

$\vec{\boldsymbol{Z}}_{\boldsymbol{s}}[\boldsymbol{t}]$ is the current best solution guiding others.

Step 5: Update Velocity and Parameters

Adjust velocities $\vec{\boldsymbol{V}}_{\boldsymbol{i}}[\boldsymbol{t}]$ and dynamic control parameter $\boldsymbol{\mu(t)}$ based on the iteration progress, to shift from exploration to exploitation.

Step 6: Termination Check

If the maximum iteration $\boldsymbol{T}$ is reached or the fitness value converges, terminate the loop.

Else, increment $\boldsymbol{t}$ and return to Step 2.

Step 7: Output the Result

Select the best candidate solution.

Classify the kidney stone into one of the predefined size ranges based on the final position of $\vec{\boldsymbol{Z}}_{\boldsymbol{s}}[\boldsymbol{t}]$.

End

Moreover, algorithm 2 initiates by generating a diverse population of candidate solutions representing possible KS size classifications. Each candidate's performance is evaluated using a fitness function that balances exploration and exploitation through eccentricity energy, normalized distances, and dynamic scaling factors. By normalizing the feature distances and computing a global fitness score, the algorithm identifies important regions in the search space. Candidates are then updated iteratively, with their positions influenced by the best solution found so far, adjusted using random perturbations to avoid premature convergence. As iterations progress, ISKOA refines the search by controlling velocity and diversity, ensuring precise classification of kidney stone sizes (<5 mm, 5-10 mm, 10-20 mm, >20 mm). This dynamic, adaptive optimization analyzes augmented data from AP-GAN for effective handling of distorted ultrasound features, leading to accurate and reliable KS size estimation. The output from this classification stage seamlessly transitions into the next block of the image retriever, which focuses on mitigating annotator bias and enhancing the accuracy of the image.

2.3.1 Unified hierarchical image retriever

The final output layer of the network introduces the Unified Hierarchical Image Retriever, which is pivotal in addressing the often-overlooked issue of annotator bias, which is an error that arises when human interpretations influence the labeling of complex medical imagery. This component is specifically designed to ensure that the retrieval process remains objective and finely tuned to capture the most relevant and discriminative features of KS images. To achieve this, a Generative Retrieval Method is incorporated, which bridges the gap between generated and real data using a unified decision-making module. This module autonomously determines the best matches by comparing both the synthetically generated and the retrieved image features, ensuring that only the most accurate representations of the classified KS are considered. This unification not only increases the precision of retrieval but also minimizes misinterpretations that typically result from manual annotations, ensuring that critical diagnostic cues are not missed.

To further enhance the retrieval accuracy and suppress the consequences of noisy labeling, the system integrates a Hierarchical Image Retrieval Technique (HIRT). This dual-stage approach works first by creating a reference index of high-level visual feature vectors from the training set. In the second stage, a detailed pixel-level comparison is conducted between the query image and each candidate, identifying the class correspondence for every pixel. This granular evaluation allows the retriever to discern both semantic and visual similarities, thereby refining the matching process. The core engine behind this comparison uses the Manhattan distance metric, which calculates the direct discrepancy between the classified query and the candidate images, reducing ambiguity and enhancing retrieval robustness. This intelligent mechanism not only corrects labeling inconsistencies but also directly supports clinical decision-making by ensuring that the retrieved images offer reliable comparisons and forms a solid foundation for further validation and therapeutic planning.

3 Results and Discussion

This part explores the experimental results in detail to comprehend the usefulness of the AS-KY-UIR framework. By methodically examining important performance measures, critical insights into how the proposed modifications contribute to enhanced detection, classification accuracy, and retrieval efficiency. These results demonstrate the architectural design's potential for practical medical imaging scenarios in addition to validating it.

3.1 Description of the dataset

The dataset[33] utilized in this study consists of 9,416 ultrasound kidney images, meticulously curated to support robust KS detection. It includes 4,414 normal and 5,002 KS images, captured across multiple scan centres and hospitals under strict compliance with patient privacy protocols. To ensure diversity in anatomical representation and scanning conditions, the dataset incorporates images acquired using a variety of ultrasound systems, including SAMSUNG RS85, HS60, RS80A, and HS70A. This heterogeneity not only enhances the generalizability of the AS-KY-UIR framework but also reflects real-world clinical variability, making it a reliable benchmark for evaluating augmentation, segmentation, classification, and retrieval tasks involved in automated KS diagnosis.

3.2 Experimental system configuration

The software and computational configuration required to implement and assess AS-KY-UIR is described in this section. Every configuration

element, including device specs, library versions, and processing settings, is described to guarantee the accuracy and consistency of the outcomes. This creates a strong basis for the controlled and transparent interpretation of performance measurements. Table 1 displays the system configuration for the AS-KY-UIR model.

3.3 Stimulation results of AS-KY-UIR

The simulation results of the AS-KY-UIR are shown in this section 4.3, with a focus on how well it performs at several important stages, including augmentation, segmentation, classification, and retrieval. The AS-KY-UIR's clinical dependability and robustness are demonstrated by evaluating the results using both quantitative performance indicators and qualitative visual comparisons, while the replication of AS-KY-UIR is given below.

The conversion from the original Kidney ultrasound image to its improved pre-processed form is shown in Fig. 6. The input image in Fig. 6(a) struggles from low contrast and comet tail artifacts, which make it difficult to see the anatomy. The proposed Contrast Adaptive Self-Guided Pixel Wise Filter and Histogram Equalization approach greatly improves contrast while maintaining textural and structural integrity, as seen in Fig. 6(b). Improved visibility of crucial features necessary for precise segmentation and classification in subsequent tasks is ensured by this pre-processing step, where further simulated images of AS-KY-UIR is given in Fig. 7.

Figure 7 presents the progressive simulation of the AS-KY-UIR model for three critical stages. In Fig. 7(a), the enhanced image demonstrates refined edge details and improved anatomical clarity, achieved through the Panoptic pipeline. Figure 7(b) highlights the isolated kidney stone regions, effectively segmented from surrounding tissues using the Panoptic CSWin Transformer and analysed by the ISKOA YOLOv8n network. Figure 7(c) depicts the output of the Unified HIRT, showcasing a retrieved image that closely matches the query in both spatial and compositional attributes. This retrieval accounts for both semantic relevance and pixel-wise similarity. Together, the figures reflect the model's robust ability to preserve detail, accurately detect KS features, and retrieve clinically relevant images.

3.4 Assessment of the AS-KY-UIR performance

An analysis of the AS-KY-UIR model with an emphasis on performance is provided in this section. Using crucial evaluation metrics, the AS-KY-UIR system's accuracy, precision, recall, and F1-score are statistically evaluated in the figures given below.

The classifier accuracy progression over training epochs for the proposed AS-KY-UIR framework is symbolized in Fig. 8. Starting at approximately 87.5% in the initial epoch, the model demonstrates a consistent and significant improvement in accuracy, reaching above 98% by the 25th epoch, and results in 99.64%. This steady increase indicates the effective

Table 1 — System configuration.

Components	Description
OS	Windows 10 Professional 64 bit
RAM	8GB
Processer	intel(R) Core (TM) i3-4130 CPU @ 3.40GHz 3.40 GHz
Tool	Python

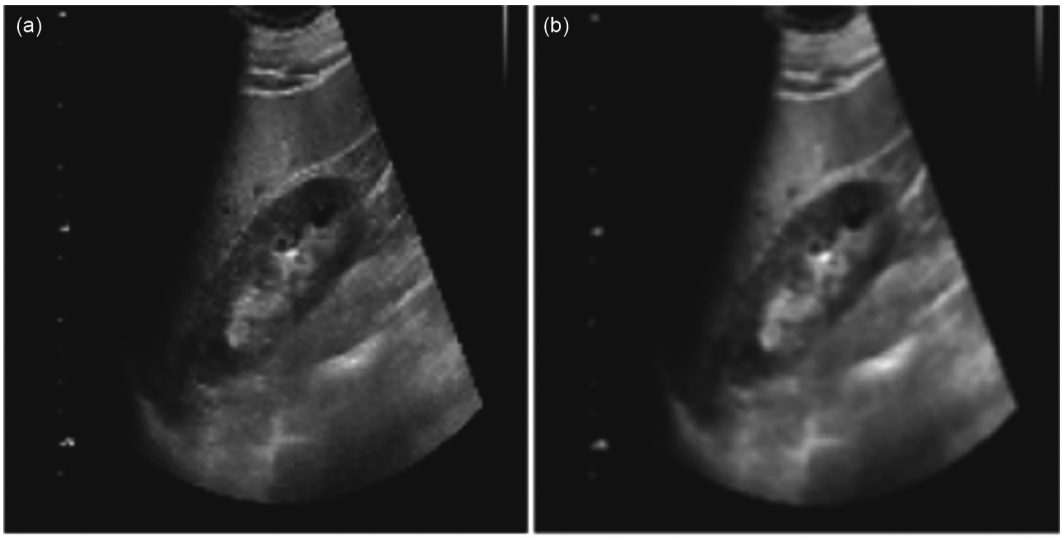


Fig. 6 — Analysis of (a) Input and (b) Pre-Processed kidney ultrasound images.

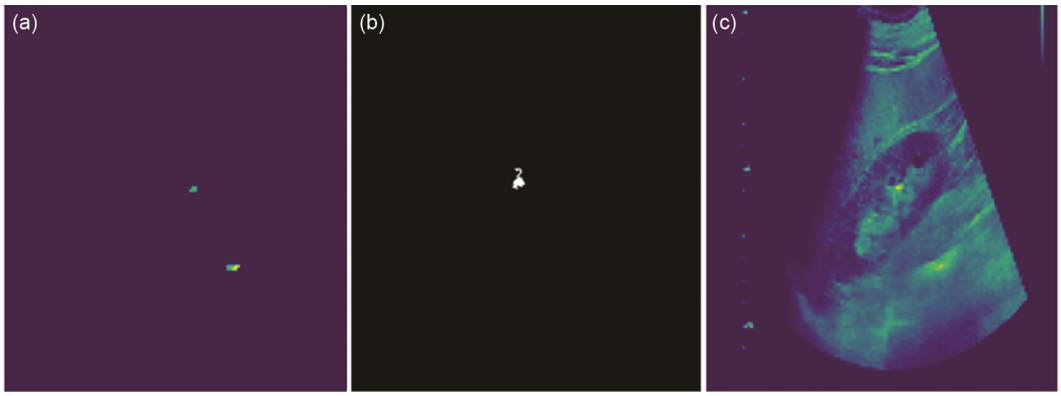


Fig. 7 — Examination of (a) Feature enhancement, (b) Stone isolation and (c) Unified retrieval representation of kidney ultrasound images.

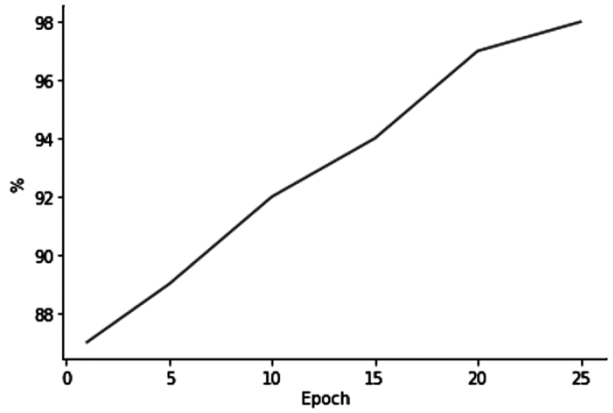


Fig. 8 — Classifier accuracy trend across training epochs in AS-KY-UIR model.

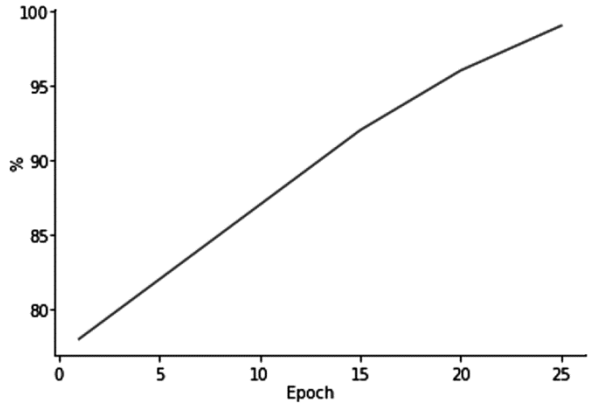


Fig. 9 — Progressive enhancement of classifier precision in AS-KY-UIR model.

learning and optimization capacity of the AS-KY-UIR pipeline, driven by its integrated components such as AP-GAN augmentation, AF-PT-SN segmentation, and ISKOA classification. The sharp rise in accuracy validates the robustness and adaptability of the model in accurately identifying kidney stone patterns from ultrasound imagery.

The graph in Fig. 9 describes the steady and significant enhancement in classifier precision achieved by the AS-KY-UIR framework throughout 25 training epochs. At the initial epoch, the precision starts at approximately 78%, reflecting the AS-KY-UIR's early-stage learning capability. As training progresses, the precision experiences a consistent upward trajectory of crossing 85% by the 10th epoch, 91% at epoch 15, and eventually reaching close to 99.58% by the 25th epoch. This marked improvement showcases the model's increasing ability to correctly identify true positives while minimizing false positives.

The line graph intensely captures the ascending trajectory of the recall rate for the AS-KY-UIR model over a series of 25 training epochs, shown in Fig. 10. The AS-KY-UIR demonstrates early sensitivity in detecting true positive cases of kidney stones, starting at about 84%. By the fifth epoch, recall increased to 87%, demonstrating the framework's increasing ability to reduce false negatives. At epoch 15, this momentum reaches 91%, and by epoch 25, it reaches above 97%. Recall consistently rising over time, resulting in 99.52%, which is a sign of the AS-KY-UIR's improved ability to retrieve pertinent positive examples, which implies both better detection and strong learning behavior in response to training.

Figure 11 illustrates the progressive amplification of the F1-Score achieved by the AS-KY-UIR framework over 25 epochs of training. Commencing at a baseline performance of 81% in the initial epoch, the model quickly ascends to 86% by the 5th epoch, signalling a balanced improvement in both precision and recall. Although the growth rate slightly tapers around epoch 10 at 87%, the model regains momentum and reaches a significant milestone of 92% at epoch 15 by synergistic coordination between

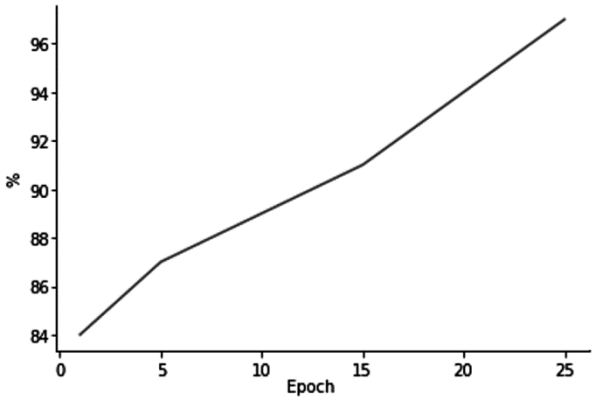


Fig. 10 — Epoch-wise evolution of classifier recall in AS-KY-UIR framework.

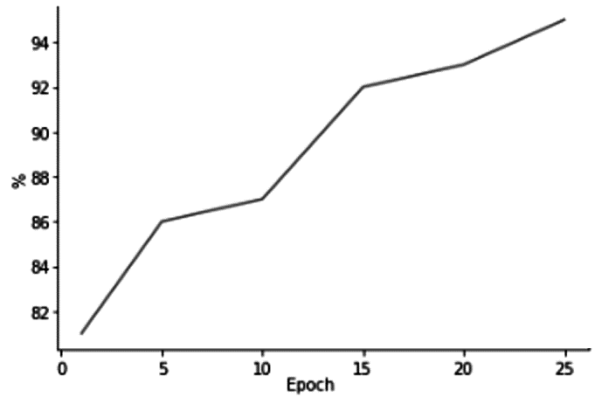


Fig. 11 — Progressive enhancement of classifier F1-score across epochs in AS-KY-UIR architecture.

the AF-PT-SN segmentation and the ISKOA classifier. By the 25th epoch, the F1-Score reaches above 95%, resulting in 99.6%, which underscores the AS-KY-UIR's evolving competence in harmonizing sensitivity and specificity for robust kidney stone detection and retrieval.

3.5 Collation of AS-KY-UIR with alternative methodologies

The effectiveness of the AS-KY-UIR model is further confirmed in this section by a comparison with other efficiently used approaches like Densenet121, MobileNetV3, GoogleNet, Ensemble model[27], Support Vector Classifier (SVC), K-Nearest Neighbors (KNN), Random Forest (RF), and Decision Tree (DT)[24]. The comparison of the performance results with metrics includes accuracy, precision, recall, F1-score, sensitivity, specificity, Dice Coefficient, and Area under the curve (AUC). These results draw attention to the advantages, disadvantages, and special qualities of AS-KY-UIR, providing a thorough grasp of its improvements over current methods in the fields of kidney stone

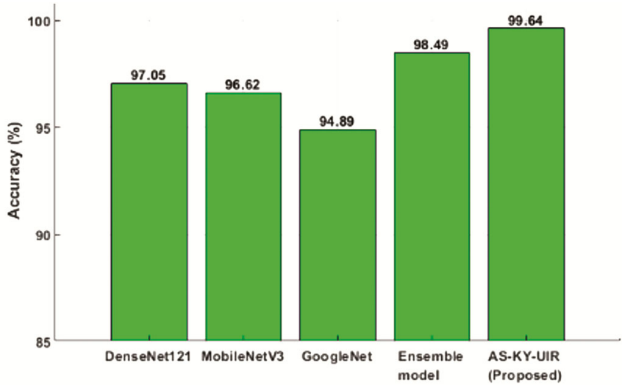


Fig. 12 — Accuracy (%) comparison across baseline models and the AS-KY-UIR model.

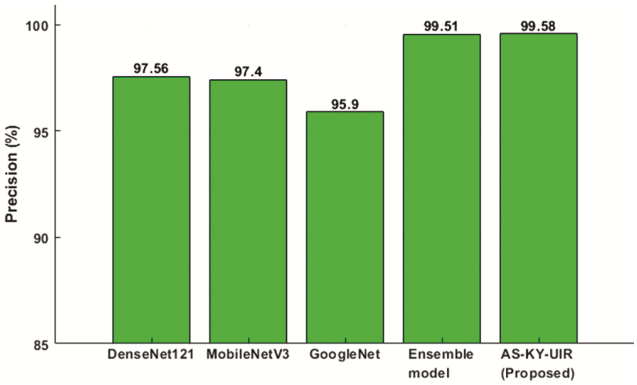


Fig. 13 — Precision analysis of AS-KY-UIR model compared to established architectures.

diagnosis and image retrieval, which is represented by the performance below.

This bar graph in Fig. 12 delineates the accuracy achieved by various deep learning models, highlighting the progressive improvements from traditional architectures to the proposed framework. Starting with foundational models, GoogleNet records a modest accuracy of 94.89%, while MobileNetV3 and DenseNet121 exhibit incremental gains of 96.62% and 97.05%, respectively. The Ensemble model further boosts performance, reaching an accuracy of 98.49% through feature-level synergy. Ultimately, the AS-KY-UIR model emerges with a superior accuracy of 99.64%, reflecting its capacity to learn and generalize over diverse input conditions effectively. This heightened accuracy stems from the integration of feature extraction and amplified discriminative learning across multi-scale layers of a unified network and optimization for a remarkably precise classification system.

This precision comparison in Fig. 13 reveals a striking edge held by the AS-KY-UIR model, which finalizes the highest precision rate of 99.58%,

underscoring its sharp accuracy in minimizing false positives during kidney stone detection. DenseNet121 and MobileNetV3 both attain strong scores of 97.56% and 97.4%, respectively, but the ensemble model closes the gap at 99.51%. However, GoogleNet drops to 95.9%, exposing its relative shortcomings. The Unified Hierarchical Image Retriever and the Improved SqueezeNet with Kepler-optimized YOLOv8n backbone are strongly connected, which contributes to AS-KY-UIR's performance. Thus, AS-KY-UIR accurately separates actual stone areas from ultrasound images while removing noise and extraneous structures.

The AS-KY-UIR framework leads with an amazing recall of 99.52%, demonstrating its exceptional sensitivity in correctly detecting all important KS locations as provided in Fig. 14. The bar chart provides a clear comparison of recall rates across different segmentation backbones. GoogleNet lags somewhat at 95.72%, while rival models like DenseNet121 and MobileNetV3 achieve up to 97.1% and 96.92%, respectively. An ensemble model even achieves 97.69%. The AS-KY-UIR pipeline's integration of the AP-GAN for Augmentation with the AF-PT-SN-based Segmentation module is directly responsible for this increased recall. By reducing false negatives and guaranteeing that no important diagnostic feature is overlooked, this synergy enables the AS-KY-UIR model to reliably identify stone borders under a variety of imaging settings.

The AS-KY-UIR model achieved the greatest F1-score of 99.6% in a performance that effectively strikes a balance between recall and precision in ultrasound-based kidney stone identification, as depicted in Fig. 15. While DenseNet121 and MobileNetV3 lag significantly behind with 98.3% and 97.83%, respectively, the ensemble model is not far behind at 98.5%. While steady, GoogleNet's somewhat lower 96.73% performance suggests that precision or recall may occasionally be sacrificed. The integration of the Kepler-optimized YOLOv8n module with the Adaptive CSWin-based SqueezeNet architecture produced an exceptional result for AS-KY-UIR. To ensure that the model could intelligently traverse through chaotic ultrasound textures while preserving clinical accuracy at scale, this strategic convergence allowed for robust feature synergy.

Figure 16 symbolizes a sensitivity performance evaluation where the model's ability to correctly detect positive cases is scrutinized across five architectures. GoogleNet reaches only 96.24%, trailing behind the rest, while MobileNetV3 and DenseNet121 post improved sensitivities of 97.23% and 98.05%, respectively. Impressively, the Ensemble model climbs to 98.7%, demonstrating the value of multi-architecture fusion. Yet, it is the AS-KY-UIR model that achieves the best, attaining a leading 99.19% sensitivity, marking its robustness in capturing true positives with minimal omission. This high-performance edge is achieved by the integration

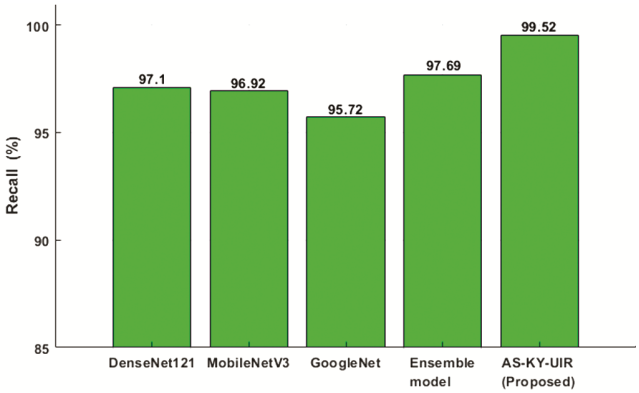


Fig. 14 — Recall performance evaluation of AS-KY-UIR against benchmark models.

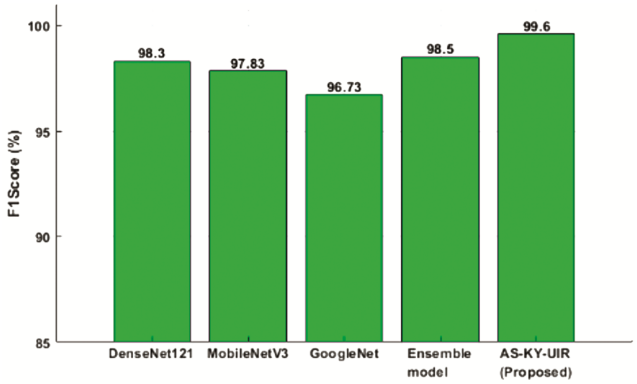


Fig. 15 — F1-score comparison of AS-KY-UIR model with existing architectures.

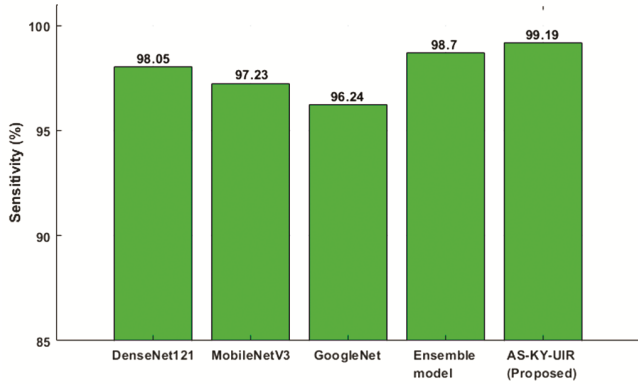


Fig. 16 — Sensitivity (%) analysis of competing models in stone type classification.

of an Attention Module and the transformer block, which together empower the AS-KY-UIR model to capture both local and long-range dependencies in the input feature maps, yielding near-perfect sensitivity in complex stone image classification tasks.

In this bar chart in Fig. 17, a compelling performance contrast is drawn between conventional CNN architectures and the AS-KY-UIR model. Notably, GoogleNet lags with a specificity of 95.8%, while MobileNetV3 and DenseNet121 deliver moderate gains of 97.32% and 97.93%, respectively. The Ensemble model, which combines multiple CNN backbones steps ahead with 98.21% showcasing the benefits of hybrid learning. However, the spotlight is firmly on the AS-KY-UIR framework, which pushes the boundary further, achieving a specificity of 99.27%, a clear leap that underscores the superior discriminatory power of the model. This remarkable specificity is attributed to the integration of the Prewitt Wavelet Transform in the input phase and the Improved Relief Feature Selector in the backbone-neck transition. Together, these modules finely sculpt the feature space, enabling AS-KY-UIR to outperform all baselines in false-positive minimization.

The segmentation of the proposed AS-KY-UIR model is provided in Fig. 18, which achieves a Dice coefficient of 99.21%, surpassing leading architectures such as DenseNet121 score of 99.33%, MobileNetV3 of 96.18%, GoogleNet of 94.71%, and even an Ensemble model of 98.82%. The clear dominance of AS-KY-UIR in this metric underscores its precision in delineating complex boundaries in ultrasound imagery. This remarkable performance is made possible by the integration of the AF-PT-SN for refined feature, coupled with the Improved Kepler YOLOv8n-based segmentation pipeline, ensuring that even the most intricate stone regions are captured with high fidelity.

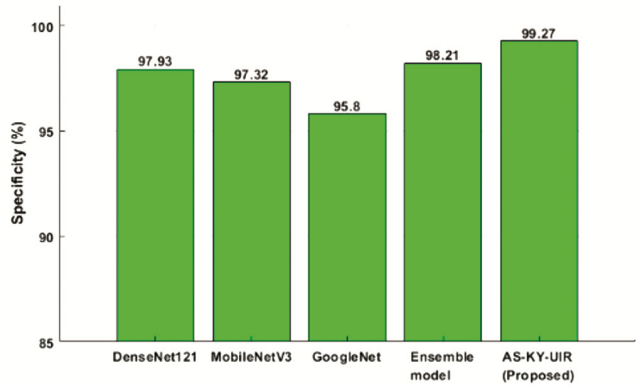


Fig. 17 — Comparison of specificity (%) among different models for stone classification.

Illustrated in a bar chart in Fig. 19 is the AUC (%) performance comparison between classical classifiers SVC, KNN, RF, and DT, against the proposed AS-KY-UIR framework, which triumphantly secures the highest AUC at 99.67%. The prevailing models achieved 94.2%, 97.6%, 98.5%, and 94.1%. This significant leap in detection capability demonstrates the strength of our hybridized approach. The superior performance of AS-KY-UIR is largely attributed to the integration of the Adaptive CSWin SqueezeNet feature extractor and the Unified Image Retrieval mechanism, which together amplify contextual precision and deep feature alignment. This robust synergy ensures heightened classification accuracy, especially in complex ultrasound imagery scenarios.

3.6 Discussion of AS-KY-UIR framework

The section highlights the effectiveness of the proposed model in addressing the key challenges of kidney stone detection through ultrasound imaging. By integrating AP-GAN for augmentation, AF-PT-SN for segmentation, and Improved Kepler YOLOv8n for

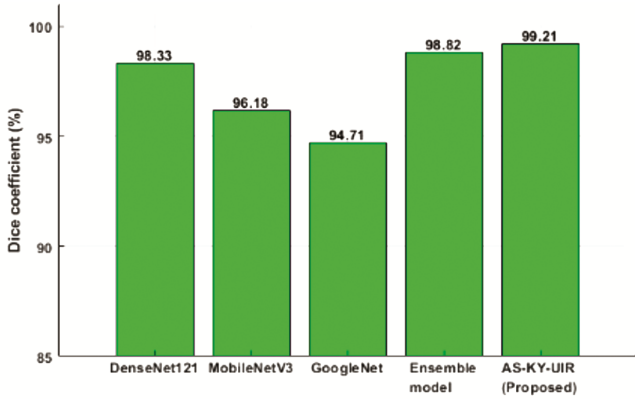


Fig. 18 — Dice coefficient comparison of AS-KY-UIR with benchmark segmentation models.

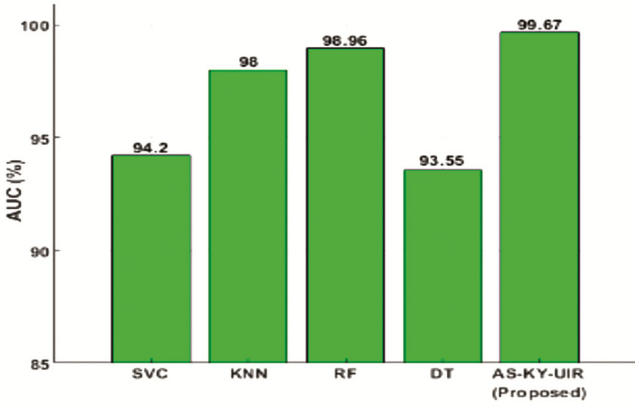


Fig. 19 — Comparative AUC analysis of AS-KY-UIR against traditional classifiers.

classification and retrieval, the framework ensures enhanced image quality, precise feature localization, and accurate classification. It also emphasizes how the model mitigates issues such as domain shift, closure loss, and annotator bias, resulting in improved diagnostic performance and reduced false positives.

The AS-KY-UIR model delivers a remarkable leap in performance compared to both traditional deep learning architectures and classical machine learning classifiers. While GoogleNet, MobileNetV3, and DenseNet121 achieve competitive accuracies of 94.89%, 96.62%, and 97.05% respectively, AS-KY-UIR surpasses them with a striking 99.64% accuracy. Precision climbs to 99.58%, clearly outpacing DenseNet121 and MobileNetV3, which linger around 97.5%. Recall is equally impressive at 99.52%, ahead of ensemble models and far above GoogleNet's 95.72%. The model further shines in sensitivity and specificity, scoring 99.19 and 99.27 percent, respectively, compared to the ensemble model's 98.7 and 98.21%. In segmentation, the Dice coefficient reaches an outstanding 99.21%, outperforming all baselines including MobileNetV3 and even ensemble models. Lastly, with an AUC of 99.67 percent, AS-KY-UIR leaves behind traditional classifiers such as SVC, KNN, RF, and DT, reinforcing its superior diagnostic capability. These achievements are the result of an intelligently fused architecture combining advanced segmentation, feature enhancement, and retrieval mechanisms tailored for kidney stone detection.

4 Conclusion

In conclusion, this study is developed to address the issues with accurate KS segmentation, classification, and reliable retrieval of KS from ultrasound images with the aid of an adaptive CSWin SqueezeNet-based improved kepler YOLOv8n model within a unified image retrieval (AS-KY-UIR). To achieve this, the framework merges four essential components: Artificial Filet-Panoptic CSWin SqueezeNet (AF-PT-SN) to provide more effective KS segmentations from complex structures, Improved Kepler optimized Prewitt Relief YOLOv8n for precise classification, and a unified hierarchical image retriever (HIRT) to decrease annotator bias while improving the similarity matching. The effectiveness of the architecture is validated via simulation on ultrasound kidney image datasets and demonstrates that the proposed AS-KY-UIR enhances segmentation accuracy, reliable classification, and retrieval performance compared to conventional methods, along with higher robustness and finer feature separation. Thus, the framework finds a solution to effectively address the problem of traditional approaches in complex KS analysis, enabling a more accurate, reliable and efficient solution for ultrasound-based KS detection band retrieval. Future work will focus on integrating with explainable AI, federated learning, cross-modality, real-time implementation, larger annotated datasets and flexibility across various domain techniques with high clinical applicability. "In addition, the addition of 3D ultrasound, multi-modal imaging, patient history, and multi-institutional training assists in more personalized and context-sensitive approaches, contributing towards progress in precision urology.

References

1 Aiumtrakul N, Thongprayoon C, Suppadungsuk S, Krisanapan P, Pinthusopon P, Mao M A, *et al.*, *Clin Pract*, 14 (3) (2024) 915.

2 Gulhane M, Kumar S, Choudhary S, Rakesh N, Zhu Y, Kaur M, *et al.*, *SLAS Technol*, 29 (4) (2024) 100159.

3 Akram M, Jahrreiss V, Skolarikos A, Geraghty R, Tzelves L, Emilliani E, *et al.*, *J Clin Med*, 13 (4) (2024) 1114.

4 Lopez-Tiro F, Estrade V, Hubert J, Flores-Araiza D, Gonzalez-Mendoza M, Ochoa G, *et al.*, *IEEE Access*, "12 (2024) 10736.

5 Chaki J & Uçar A, *IEEE Access*, 12 (2024) 32894.

6 Kilic U, Karabey Aksakalli I, Tumuklu Ozyer G, Aksakalli T, Ozyer B & Adanur S, *Int J Intell Syst*, (2023) 3801485.

7 Neha F, *7th International Conference on Computing, Communication, Control and Automation (ICCUBEA)* (IEEE), (2023) 1.

8 Huang Z H, Liu Y Y, Wu W J & Huang K W, *Bioengineering*, 10 (2023) 970.

9 Pande S D & Agarwal R, *IEEE Access*, 12 (2024) 21147.

10 Chew B H, Wong V K, Halawani A, Lee S, Baek S, Kang H, *et al.*, *Urolithiasis*, 51(1) (2023) 117.

11 Yan C & Razmjooy N, *Biomed Signal Process Control*, 86 (2023) 104951.

12 Narasimhulu C V, *IET Image Process*, 15(9) (2021) 2080.

13 Preedanan W, Suzuki K, Kondo T, Kobayashi M, Tanaka H, Ishioka J, *et al.*, *IEEE Access*, 11 (2023) 25702.

14 Sharen H, Narendra M & Anbarasi L J, *IEEE Access*, (2024).

15 Ibragimov B, Zhen J & Ayvali E, *IEEE Access*, 11 (2023) 90358.

16 Shamna P, Govindan V K & Nazeer K A, *J King Saud Univ Comput Inf Sci*, 34(2) (2022) 58.

17 Jiang D & Kim J, *Appl Sci*, 11(12) (2021) 5701.

18 Alexa R, Kranz J, Kramann R, Kuppe C, Sanyal R, Hayat S, *et al.*, *Eur Urol Open Sci*, 62 (2024) 19.

19 Bhandari M, Yogarajah P, Kavitha M S & Condell J, *Appl Sci*, 13 (2023) 3125.

20 Mousavi A, Thaker K N, Ackerman J E, Diaz N, Martin R, Tipton CD, *et al.*, *Pathogens*, 13(11) (2024) 942.

21 Yousif S M A, Halawani H T, Amoudi G, Birkea F M O, Almunajam A M & Elhag A A, *Alex Eng J*, 100 (2024) 220.

22 Alsekait D M, Saleh H, Gabralla L A, Alnowaiser K, El-Sappagh S, Sahal R, *et al.*, *Appl Sci*, 13(6) (2023) 3937.

23 Tirupal T, Reddy V S C, Manoj C, Venumadhav S, Mahesh A & Rakesh B, *Int J Innov Multidiscip Sci Res*, (2023).
24 Asaye Y A, Annamalai P & Ayalew L G, *Sci Afr*, 28 (2025) e02618.
25 Jadhav A N, Sharathamani T, Souza M D, Thiyagarajan A, Rokade M D, Shelke R M, Patil A R & Swapna S, *J Neonatal Surg*, 14 (2025).
26 Balanageshwara S & Aashitha S, *Cuest Fisioter*, 53(03) (2024) 1153.
27 Mahalakshmi S D, *Comput Syst Sci Eng*, 44(2) (2023).
28 Wang X, Snášel V, Mirjalili S, Pan JS, Kong L & Shehadeh HA, *Knowl-Based Syst*, 295 (2024) 111737.
29 Shen Z, Wang C, Li F, Liang J, Li X & Qu D, *Appl Sci*, 14 (23) (2024) 11033.
30 Liu X, Gao P, Yu T, Wang F & Yuan RY, *Inf Fusion*, 113 (2025) 102634.
31 Tenekeci ME, Abdulazeez ST, Karadağ K & Modanli M, *Multimed Tools Appl*, (2024) 1.
32 Qian Z, Zhang Y, Pu D, Xie G, Pu D & Ye M, *Mathematics*, 13 (2025) 405.
33 Gurjeetkaurmangat, *Kidney Ultrasound Images—Stone and No Stone*, Kaggle dataset, available at *https://www.kaggle.com/datasets/gurjeetkaurmangat/kidney-ultrasound-images-stone-and-no-stone.*