International Journal of Engineering Trends and Technology
ISSN: 2231-5381/ https://doi.org/10.14445/22315381/IJETT-V74I11P121

Volume 74 Issue 11, 304-328, November 2026
© 2026 Seventh Sense Research Group®

Original Article

DeepSurv - Assisted Survival Modeling using Regularized Filtered Back Projection (FBP) - Vision Transformer (ViT) from Computed Tomography (CT), Magnetic Resonance Imaging (MRI) and Ultrasound Imaging for Renal Disease Prediction

Scinthia Clarinda S[1], Sonia Jenifer Rayen[2]

[1,2]*Department of Computer Science and Engineering, School of Computing, Sathyabama Institute of Science and Technology, Chennai, Tamilnadu, India.*

[1]*Corresponding Author: scinthia9508211@outlook.com*

Received: 08 January 2026 Revised: 22 July 2026 Accepted: 07 August 2026 Published: 25 November 2026

Abstract *- Renal disorders are a main cause of kidney failure and remain a pressing public health concern. Conventional diagnosis largely depends on nephrologists' subjective, labor-intensive, and variable expertise in assessing renal architecture. To overcome these limitations, this study proposes a hybrid framework, DeepSurv-FBP-ViT, which integrates advanced image reconstruction, transformer-based representation learning, and deep survival modeling for accurate prediction of renal disease progression across multiple imaging modalities. Regularized Filtered Back Projection (FBP) reconstruction is applied to CT data to suppress noise and preserve anatomical details, while standardized preprocessing pipelines are employed for MRI and ultrasound images to ensure cross-modality consistency. A Vision Transformer (ViT) is used to analyze the preprocessed and reconstructed images by capturing both local structural features and global contextual relationships across modalities through self-attention. The imaging-derived features are then combined with relevant clinical variables and processed using DeepSurv, a deep learning model for survival analysis, to estimate patient-specific risk scores and survival probabilities. Evaluation on publicly available renal imaging datasets demonstrates that the proposed DeepSurv-FBP-ViT multimodal pipeline outperforms traditional Cox models in terms of both concordance index and risk stratification. By leveraging information from CT, MRI, and ultrasound, this approach provides a comprehensive, non-invasive, and reliable tool for renal disease prognosis and clinical decision support.*

Keywords *- DeepSurv, Filtered Back Projection (FBP), Renal disease prediction, Survival analysis, Vision Transformer (ViT).*

1. Introduction

The kidneys are among the most structurally and functionally complex organs in the human body, playing a vital role in maintaining physiological homeostasis. Anatomically, each kidney consists of an outer renal cortex and an inner renal medulla containing multiple cone-shaped structures known as renal pyramids. These structures, together with the nephrons and collecting ducts, facilitate blood filtration, urine formation, and the regulation of fluid and electrolyte balance [1]. In healthy adults, each kidney typically weighs approximately 120-170 g and measures about 10-12 cm in length, 5-7 cm in width, and 2.5-3 cm in thickness, although these dimensions vary depending on age, sex, and body size. The kidneys perform several essential physiological functions, including the elimination of metabolic waste products, maintenance of water and electrolyte balance, regulation of acid-base equilibrium, blood pressure control, and endocrine activities such as erythropoietin secretion, renin release, and vitamin D activation [2]. Impairment of renal function can significantly disrupt metabolic regulation, cardiovascular stability, electrolyte balance, and endocrine homeostasis, potentially contributing to chronic kidney disease, cardiovascular complications, and an increased risk of mortality [3]. Renal Diseases (RDs) impact more than 700 million people globally and are responsible for nearly 1.2 million fatalities annually, making them a serious global public health concern. The mortality rates due to renal failure rose by almost 38% in India from 2003 to 2013. This increasing prevalence of diabetes, obesity, hypertension and ageing populations are continuing to increase the burden of kidney diseases [4]. Kidney cysts,

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

stones, and tumors are the most common renal abnormalities. Kidney stones are hard nodules of mineral that can build up in the urinary tract and be extremely painful and obstruct urine flow [5].

Tumors of the kidney are generally cancerous and, in particular, Renal Cell Carcinoma (RCC) is considered to be one of the most serious forms of cancer and should be diagnosed and treated early for best prognosis. In a similar manner, Polycystic Kidney Disease (PKD) is a disease that involves the growth of many fluid-filled cysts which eventually begin to damage the ability of the kidneys to function properly, and more than half of those with the disease eventually develop renal failure [6].

Early and accurate diagnosis becomes essential for developed treatment planning, lesser disease progression and developed patient survival. Medical imaging is a significant part of the diagnosis and evaluation of renal disorders. Kidney illnesses are frequently diagnosed, assessed, and tracked using clinical imaging modalities such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Ultrasound (US), and X-ray. CT imaging is one of the most useful modalities for detecting renal cysts, stones, renal tumors and other structural abnormalities because of its excellent soft-tissue visualization and high spatial resolution and detailed 3-D anatomical information [7].

Figure 1 illustrates representative Computed Tomography (CT) images of normal kidneys and kidneys affected by various renal disorders, including cysts, stones, and tumors, in both axial and coronal perspectives. Renal cysts (Figures 1(a) and 1(b), water-filled areas in the kidneys). The hardened mineral deposits are known as kidney stones and are shown in Figures 1(c) and 1(d). Figures 1(e) and 1(f) exhibit kidney tumor pictures that predominantly highlight aberrant tissue proliferation within the renal structure. Normal images of the kidneys with healthy features are displayed in Figures 1 (g) and 1(h) where there are no visible abnormalities. There is also a significant difference in the size, shape, location, texture and appearance of the lesions, which reflects the complexity of automated renal disease classification.

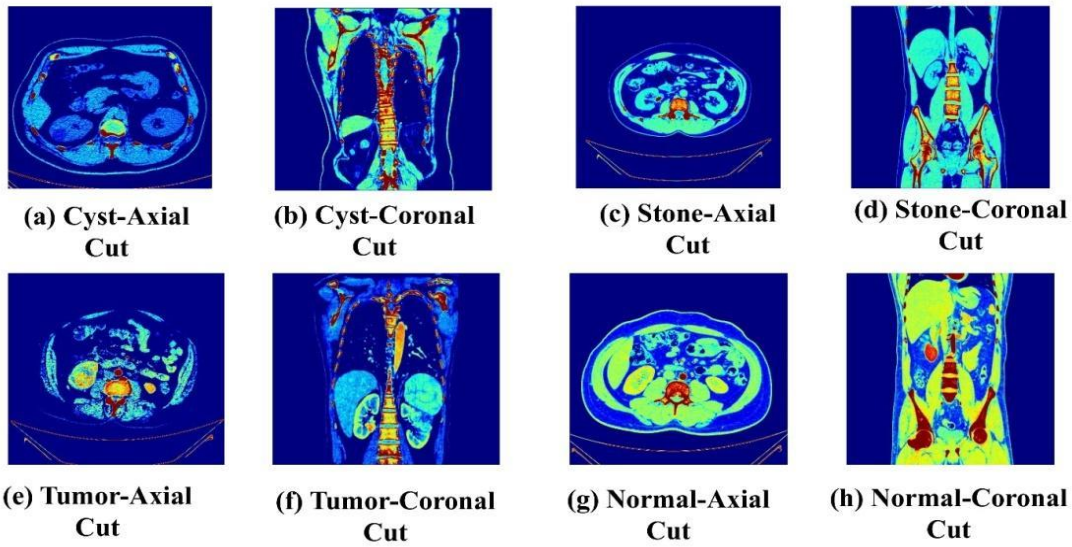


Fig. 1 An illustration of a range of kidney disorders, CT scans from the renal imaging dataset [12]

Renal diseases are becoming more common, and there is a shortage of trained nephrologists in many health care systems, which has spurred the utilize of Artificial Intelligence (AI) diagnostic tools [8]. Based on artificial intelligence, computer-aided diagnosis has shown great promise in supporting clinicians with rapid, objective, and consistent analysis of kidney Computed Tomography (CT) images. With the advent of Machine Learning (ML) and, more recently, Deep Learning (DL), which can automatically build hierarchical feature representations from medical imaging data, medical image categorisation has made tremendous strides in recent years [9-11]. Numerous studies have displayed the positive effect of CNN-based architectures for CT image classification of kidney diseases. But there are still a few drawbacks to the current approaches. Most models only use one scale of feature extraction, which is not very effective in learning local features of the lesion and global anatomical information. High-dimensional deep features may have redundant and irrelevant information, and it will make the computation more difficult and it will lead to less efficiency of classification. In addition, traditional CNN structures tend to focus on the spatial features and lack sufficient ability to represent long-range contextual information, which causes the CNN to have poor ability to discriminate between similar renal abnormalities, including cysts, tumors, and complex stones, in the same image space. Such constraints limit the strength and applicability of the current automated diagnostic systems to different clinical datasets.

1.1. Novelty and Contributions

This research introduces a novel hybrid deep learning system for automatically classifying kidney disorders from Computed Tomography (CT) images, which overcomes the limitations of existing approaches. In order to achieve the highly discriminative feature representation while suppressing redundant information, the proposed approach combines intelligent feature optimization, sequential contextual learning and multi-scale deep feature extraction in a unified framework. Compared to conventional CNN-based techniques, the suggested method provides higher classification accuracy and more computing efficiency.

It doesn't necessity big training datasets like transformer-based models necessity, and it doesn't rely on extra clinical data like multimodal fusion methods. Under stringent experimental protocols, the recommended framework was thoroughly assessed utilizing a publically accessible renal CT dataset and contrasted with cutting-edge techniques. The outcomes display how reliable it is as a computer-aided diagnosis system for the classification of multiclass renal disorders.

1.2. The Major Motivation of the Research

Renal disease progression is not uniform and is challenging to predict early on, but is a critical factor in deciding therapy, frequency of monitoring and transplant planning. Prognostic practice today relies upon imaging characteristics that may be difficult to formulate and interpret, requiring great radiologist experience, and on radiologists' ability to recognize subtle but spatially distributed features of parenchymal injury, which may differ among readers, scanners, and institutions. Survival models, such as Cox proportional hazards models, are traditionally used and have the assumption of linearity and proportional hazards, which may not be valid in renal pathology due to the presence of many nonlinear interactions between imaging and clinical variables. Furthermore, noise and streak artefacts may be amplified by traditional CT image processing pipelines, which would deteriorate image quality and negatively impact the extraction of trustworthy downstream features. Additionally, MRI and ultrasound provide modality-specific biases, thereby hampering multimodal fusion and model generalizability.

1.3. Problem Definition

This study main goal is to improve kidney illness prognosis accuracy, dependability, and computational efficiency by efficiently fusing complementary data from multimodal medical images, such as CT, MRI, and ultrasound. Most current research works are based on either a single imaging modality or on image reconstruction, deep feature extraction and survival prediction separately. These methods typically lack the ability to maintain fine anatomical structures, incorporate long-range context information into the framework, and model patient-specific survival risk in a single model. In order to tackle these problems, this study proposes the DeepSurv-FBP-ViT framework. This framework incorporates three important innovations: a Vision Transformer (ViT) for learning global and local imaging features, Regularised Filtered Back Projection (FBP) for effective reconstruction of CT images and robust suppression of noise, and multimodal data fusion for combining complementary information from CT, MRI, and US. To overcome these difficulties, this study proposes the DeepSurv-FBP-ViT framework, which contains three novel contributions, including an efficient and robust algorithm Regularized Filtered Back Projection (FBP) for noise suppressed CT image reconstruction, Vision Transformer (ViT) to learn imaging features at local and global levels, and multimodal data fusion to make use of complementary information provided by CT, MRI, and US modalities in personalized survival risk estimation. It is an integrated framework that offers an overall and clinically relevant solution to improve the prognostic accuracy and risk stratification of patients beyond the Cox proportional hazards models and single-modality deep learning approaches.

The remaining sections of this work are organised as follows. Section 2 provides an overview of important literature. Section 3 explains the recommended methodology. In Section 4, we present the results of the experiment, and then we analyse and discuss them in detail. The work is concluded, and areas for future investigation are suggested in Section 5.

2. Related Work

Utilizing clinical and demographic data, Karthikadevi and Sneha [13] examined a machine learning-based framework for predicting the risk of Chronic Kidney Disease (CKD), with the Light Gradient Boosting Machine (LightGBM) acting as the major predictive model. Since LightGBM can process large-scale data with a large number of factors, it was proved to be helpful for the risk assessment of CKD. The approach allows for early diagnosis and rapid treatment by including patient-specific clinical data like age, blood pressure, and blood test results. LightGBM fared better than other machine learning models, including Decision Trees and K-Nearest Neighbours (KNN), according to a comparison investigation. With 96% accuracy, 93% recall, and an F1-score of 89%, the model successfully predicted the existence of chronic renal disease.

Saif et al. [14] proposed a model for predicting Chronic Kidney Disease (CKD) using deep learning and deep ensemble learning. While most studies currently available focus on diagnosing whether a patient has CKD once it has occurred, their study highlighted the prediction of the occurrence of CKD in the next 6 and 12 months. While addressing significant issues such data imbalance, feature selection, and optimiser selection, the authors evaluated several deep learning models, including CNN, LSTM, and LSTM-BLSTM. Furthermore, an ensemble model was created by combining the top-performing individual models, CNN-

Adamax, LSTM-Adam, and LSTM-BLSTM-Adamax. For 6-month and 12-month CKD prediction, respectively, the suggested framework attained high prediction accuracy of 98% and 97%. However, the study mainly focused on clinical CKD prediction rather than CT image-based renal disorder classification; therefore, its applicability to automated image-based diagnosis of renal cysts, stones, and tumors remains limited.

An EOAEDL-CKDD framework for Chronic Kidney Disease Detection based on Ensemble Deep Learning was presented by Yousif et al. [15]. This framework is tuned for CKD diagnosis utilising the Eurygaster Optimisation Algorithm (EOA). The proposed approach integrates feature selection and hyperparameter tuning to develop classification performance. Initially, min-max normalization was applied to scale the dataset uniformly. Then, the most informative features were selected using the Eurygaster Optimization Algorithm (EOA).

Long Short-Term Memory (LSTM), Bidirectional Gated Recurrent Unit (BiGRU), and Bidirectional Long Short-Term Memory (BiLSTM) networks were integrated in an ensemble deep learning framework for classification in order to improve classification performance. Furthermore, the Shuffled Frog Leaping Algorithm (SFLA) was applied to optimize the hyperparameters of the ensemble classifiers. Experimental validation on a benchmark CKD dataset demonstrated that the EOAEDL-CKDD model outperformed several existing techniques in terms of detection performance.

Islam et al. [16] examined the relationship between target-class features and data variables. Predictive modeling, supported by machine learning and predictive analytics, can help develop more effective quality metrics by generating a comprehensive set of prediction models. In their study, the initial set of 25 variables, including the class label, was reduced to 30% of its original size to recognize the most significant feature subset for CKD detection. In the supervised learning task, XGBoost outperformed twelve other classifiers, achieving an accuracy of 0.983, precision of 0.98, recall of 0.98, and F1-score of 0.98.

In order to detect CKD, Yogalakshmi and Anitha [17] created a hybrid deep learning framework that combines a Support Vector Machine (SVM) classifier with a Convolutional Neural Network (CNN). In their method, the SVM functioned as the final classification model, while the CNN automatically collected high-level discriminative features from the input data. The experimental tests conducted on the CKD dataset indicated that the framework was able to classify the data with 96.18% accuracy. Moreover, data augmentation and data normalization techniques were used to further develop the robustness and generalization power. The hybrid CNN-SVM architecture showed superiority in diagnostic performance over conventional machine learning techniques, but its test was based on the classification of CKD only and did not include the classification of renal CT images in the presence of kidney cysts, stones, tumors and a normal kidney. Moreover, this model did not explicitly optimize the redundancy of features, nor explicitly consider long-range contexts, potentially reducing the efficiency of processing complex renal imaging datasets.

A hybrid CNN-SVM framework was generated by Ramu et al. [18] to develop the prediction accuracy of CKD. The CNN was utilized in the suggested method to automatically extract important characteristics, and the SVM was utilized for final classification. In a sizable clinical dataset with ten diagnostic markers, SMOTE was used to address class imbalance. With an accuracy of 96.8%, the hybrid technique outperformed the independent Random Forest and SVM models, which had respective accuracies of 94.6% and 94.8%.

Transformer-based architectures can capture global contextual information and model long-range spatial relationships; therefore, they have been widely utilized in recent studies on medical image analysis. Advanced deep learning architectures, such as Vision Transformer (ViT), Swin Transformer, and hybrid CNN-Transformer models, have demonstrated strong performance in several medical image classification tasks, particularly in the detection of renal disease.

However, although these methods improve feature representation and classification performance, they often necessitate large annotated datasets and expensive computational resources. This limits their applicability to relatively small renal CT scan datasets. Furthermore, the considerable computational complexity of many transformer-based models makes them unsuitable for real-time clinical applications.

Apart from image-based approaches, multimodal data fusion has been identified as an effective way to diagnose renal disease, combining computed tomography scans or magnetic resonance scans with clinical notes, laboratory tests and genome data. These fusion frameworks enhance the diagnostic performance by utilizing complementary information from various data sources. In reality, however, the multimodal data sources are incomplete or are not routinely collected in clinical settings, which makes the models more complex, expensive to compute and less generalizable between health care institutions.

Several recent works explored the problem of survival modeling and prognosis prediction for chronic kidney disease both from statistical and deep learning perspectives. These are methods that can estimate disease progression, renal function decline or patient survival using longitudinal clinical variables. Useful for prognosis and treatment planning, survival models are aimed mainly at predicting clinical

outcomes but not accurately distinguishing renal abnormalities (e.g. cysts, stones, tumors and normal kidney) from CT images.

2.1. Research Gap

The field of renal disease analysis has made significant strides beyond traditional CNN-based approaches, with the integration of transformer architectures, multimodal data fusion, and survival prediction models. Global contextual features and long-range spatial relationships can be captured with Vision Transformers (ViTs) and hybrid CNN-Transformer-based architectures. However, these approaches are not suitable for relatively small medical imaging datasets as they generally require large annotated datasets and significant computational power.

Likewise, multimodal data fusion techniques combining CT images with clinical, laboratory or genetic data have also demonstrated potential to improve diagnostic accuracy, nevertheless, these data are not routinely available in clinical practice and may add to the complexity of the model.

Moreover, most existing survival modeling methods concentrate on predicting patients' outcomes and the progression of the disease, and do not accurately classify renal abnormalities in diagnostic Computed Tomography (CT) images. Although these developments are made, there is no efficient framework that can be used for multi-scale feature extraction, feature optimization and contextual feature learning for multi-class renal CT image classification. Current techniques tend to include redundant feature representations, a higher number of computations and are less discriminative in distinguishing renal abnormalities sharing similar visual characteristics like cysts, stones, and tumors. Thus, there is a need for a framework which is computationally efficient, robust and improves diagnostic accuracy while enhancing the feature representation. Table 1 discusses the literature survey of proposed model.

Table 1. Literature survey for proposed method

Author	Objective	Dataset	Methods	Key Features	Performance
Karthikadevi and Sneha [13]	Predict CKD risk utilizing clinical and demographic features	CKD clinical dataset	LightGBM	Feature-based prediction using age, blood pressure, and laboratory parameters	Achieved 96% accuracy
Saif et al. [14]	Early prediction of CKD occurrence within 6 and 12 months	CKD clinical dataset	CNN, LSTM, LSTM-BLSTM, Deep Ensemble	Deep ensemble learning with optimizer selection, feature selection, and imbalance handling	Achieved 98% and 97% prediction accuracy
Yousif et al. [15]	Improve CKD diagnosis through optimized feature selection and ensemble learning	Benchmark CKD dataset	EOA + LSTM + BiGRU + BiLSTM + SFLA	Eurygaster Optimization for feature selection and SFLA-based hyperparameter optimization	Not specified numerically
Islam et al. [16]	Select best features and classify CKD	25 variables (reduced to ~30%)	12 ML classifiers, best: XGBoost	Optimized attribute subset	Accuracy: 0.983, Precision/Recall/F1: 0.98
Yogalakshmi and Anitha [17]	Develop a hybrid deep learning model for CKD detection	CKD dataset	CNN-SVM	CNN-based feature extraction with SVM classifier	Achieved 96.18% accuracy
Ramu et al. [18]	Improve CKD prediction with a mixed deep learning model	CKD clinical dataset	CNN-SVM + SMOTE	CNN feature extraction, SVM classification, and class balancing using SMOTE	Achieved 96.8% accuracy

2.2. Problem Identification for Existing Systems

- Numerous existing models focus solely on binary classification (e.g., healthy vs. diseased) and fail to accurately differentiate amongst numerous renal abnormalities, such as cysts, stones, tumors, and other structural anomalies.
- Most existing systems are trained on small-scale or non-diverse datasets that lack sufficient variation in disease stages, demographic characteristics, and imaging orientations (e.g., axial and coronal views), resulting in poor generalization and limited robustness in real-world clinical scenarios.

- The kidney's intricate structure and the subtle differences among benign and malignant features make it challenging for conventional models to extract meaningful patterns, particularly during the early stages of disease detection.
- Numerous systems rely solely on image-based data, overlooking valuable clinical and laboratory information, such as creatinine levels and patient history, that could suggestively develop prediction reliability.

3. Proposed Model

3.1. Workflow

In this section, we introduce a hybrid framework, called DeepSurv-FBP-ViT, which combines image reconstruction algorithms, transformer-based representation learning and deep survival models to predict renal disease progression across different imaging modalities accurately. The CT data is reconstructed using Regularized Filtered Back Projection (FBP) to reduce noise and highlight anatomical details, and the MRI and ultrasound images undergo standardized processing to maintain consistency across the different modalities. A novel multimodal fine-grained and global structural element extraction framework based on a Vision Transformer (ViT) with self-attention is developed to extract fine-grained local structural elements and global contextual links from multimodal pictures. In order to determine patient-specific risk scores and survival probability, the image-derived characteristics are then combined with pertinent clinical factors and examined using DeepSurv, a deep learning-based survival analysis model.

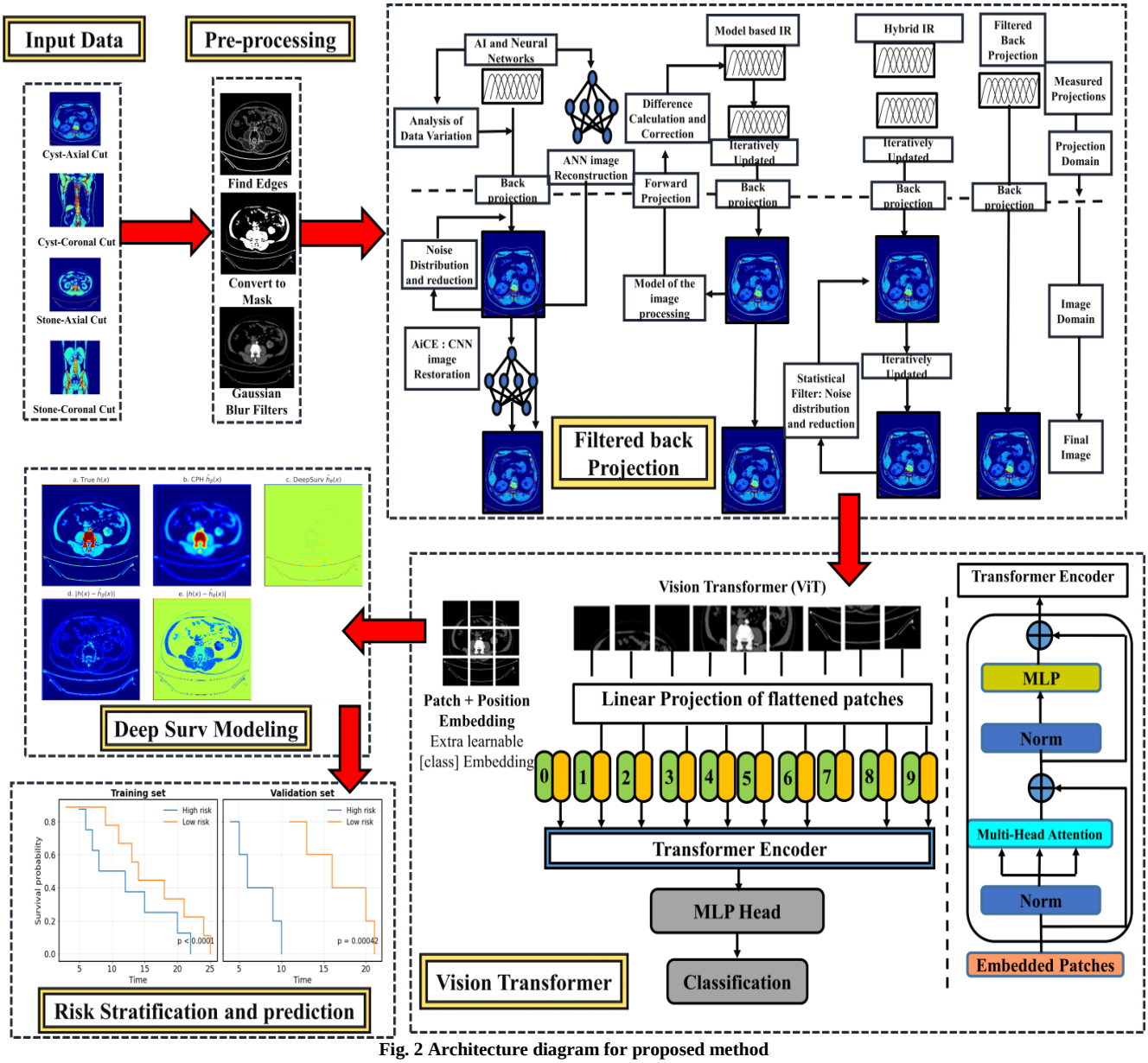


Fig. 2 Architecture diagram for proposed method

The architecture of the suggested DeepSurv-assisted survival prediction framework, which combines a Vision Transformer (ViT) and Filtered Back Projection (FBP) for precise renal disease risk prediction and survival modeling, is shown in Figure 2.

The workflow begins with input data consisting of CT scan images, including axial and coronal views of cysts and stones. These images undergo preprocessing, where edge detection, mask conversion, and Gaussian blur filtering are applied to enhance important structures and suppress noise.

The preprocessed images are then passed to the FBP module, where noise reduction, iterative reconstruction, and CNN-based image restoration are performed to recover high-quality CT images from projection data. Statistical and model-based iterative updates are incorporated to minimize artifacts and refine the reconstructed images.

Next, the reconstructed CT images are processed by the ViT module. Before being flattened, the images are separated into patches of certain sizes. Then, the positional embedding and class-token are added. Finally, the images are projected linearly into sequential representations. The subsequent Transformer encoder captures general global context information and fine-grained anatomical information through layer normalization and multi-head self-attention. Then the extracted embeddings are given to the classification heads to create a compact feature representation.

Finally, features generated by the ViT are incorporated into a DeepSurv modeling framework that combines imaging biomarkers with survival analysis to predict patient-specific outcomes. The system provides both risk stratification and prediction results such as Kaplan-Meier survival curves for training and validation cohorts. This pipeline provides accurate image reconstruction, efficient feature learning and clinically relevant risk prediction for renal disease prognosis.

3.2. Pre-Processing of CT Images

The input images are then pre-processed to enhance image quality and uniformity for use in various imaging modalities such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI) and ultrasound. The images obtained by CT are stored as Hounsfield Units (HU) and then windowed for the intensity of the images and normalized utilizing a min-max method. Normalisation and Denoising of MRI or ultrasound image.

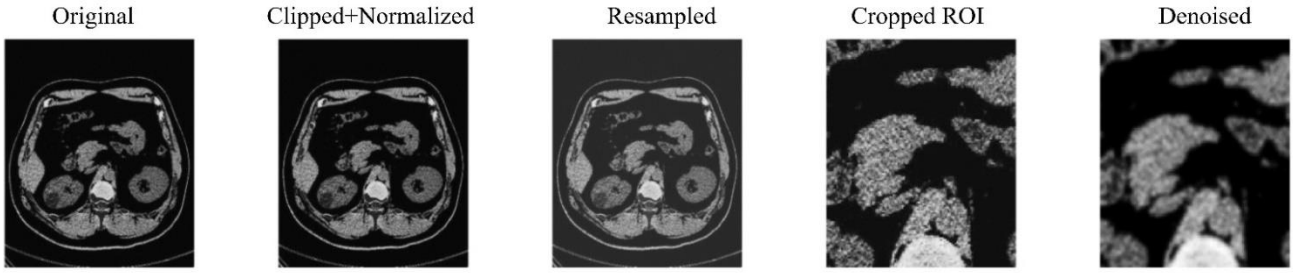


Fig. 3 Pre-Processing images for proposed technique

All images were cropped in the Region of Interest (ROI) of interest of the kidney, reduced to a fixed input size and resampled to a uniform spatial resolution. Transformations such random rotation, horizontal flipping, scaling, and translation were used to enhance the training dataset in order to increase model generalisation and lower the risk of overfitting. Figure 3 depicts the pre-processed images which are applied in the proposed technique.

The HU values were clipped to a preset range (e.g., -100 to 400 for abdominal areas) and later standardized to the [0, 1] interval using min-max scaling to maintain consistency across all samples.

$$\mathrm{I_{norm}} = \frac{\mathrm{clip(I_{HU}, HU_{min}, HU_{max}) - HU_{min}}}{\mathrm{HU_{max} - HU_{min}}} \quad (1)$$

Where I_{HU} is the original CT image in Hounsfield units, and HU_{min} and HU_{max} define the desired intensity range.

Then, the resampling is done to assure uniform voxel dimensions in all three dimensions for all the CT images. This is necessary as CT scans obtained from different sources may have different spatial resolution. Each image is interpolated (e.g., linear or cubic) to standardize the resolution of the image for subsequent analysis by adjusting the image resolution for each image according to the ratio of the voxel spacing of the image to the spacing of the target image.

$$I_{resampled}(x', y', z') = I\left(\frac{x'}{s_x}, \frac{y'}{s_y}, \frac{z'}{s_z}\right) \quad (2)$$

Where $s_x, s_y, s_z$the scale are factors for each axis, and $I_{resampled}$ is the resampled volume.

The image is cropped using either anatomical landmarks or segmentation outputs to focus the analysis on the Region of Interest (ROI), such as the kidneys. This process reduces the input size and computational overhead. The ROI is then extracted using bounding box coordinates.

$$I_{crop} = I[xmaxminmaxminmax_{min}] \quad (3)$$

Lastly, noise reduction is an optional step to reduce noise at high frequencies that might affect the model's performance. This is usually done with a Gaussian filter, for which the smoothed intensity at each voxel is given by:

$$I_{denoised}(x,y,z) = \sum_{i,j,k} I(x+i, y+j, z+k) \cdot G(i,j,k) \quad (4)$$

Here, $G(i,j,k)$signifies the weights from a 3D Gaussian kernel. Together, these preprocessing steps produce clean, uniformly scaled, and spatially focused CT images that are well suited for downstream tasks such as segmentation, classification, and survival prediction.

3.3. Regularized Filtered Back Projection (FBP) Reconstruction

For CT image reconstruction, the Regularized Filtered Back Projection (FBP) method was utilized to develop the image quality. The regularized FBP can remove reconstruction noise and streaking artifacts without compromising the anatomical structure and lesion borders compared to the conventional FBP technique. It achieves higher accuracy and robustness in the diagnostic process by providing high-quality reconstructed images that provide more reliable inputs for further feature extraction.

Regularized FBP improves the quality of reconstructed CT images by combining statistical regularization into the traditional Filtered Back Projection (FBP) method.

While conventional FBP is a computational efficient and is commonly used in clinical practice, this algorithm is sensitive to noise amplification and reconstruction artifacts, especially in low dose and sparse view CT imaging.

The traditional Filtered Back Projection (FBP) method is shown in Figure 4 [19]. To address this limitation, regularized FBP formulates image reconstruction as an optimization problem that balances fidelity to the measured projection data with a regularization term that incorporates prior knowledge, such as smoothness or sparsity.

This optimization problem is typically expressed as follows:

$$\hat{f} = arg\min_{f} \{\|Af - g\|_2^2 + \lambda R(f)\} \quad (5)$$

Where $\hat{f}$is the reconstructed image, A is the forward projection operator, g is the measured sinogram data, $R(f)$is the regularization functional, and $\lambda > 0$is a regularization parameter controlling the trade-off among data fidelity and regularization strength.

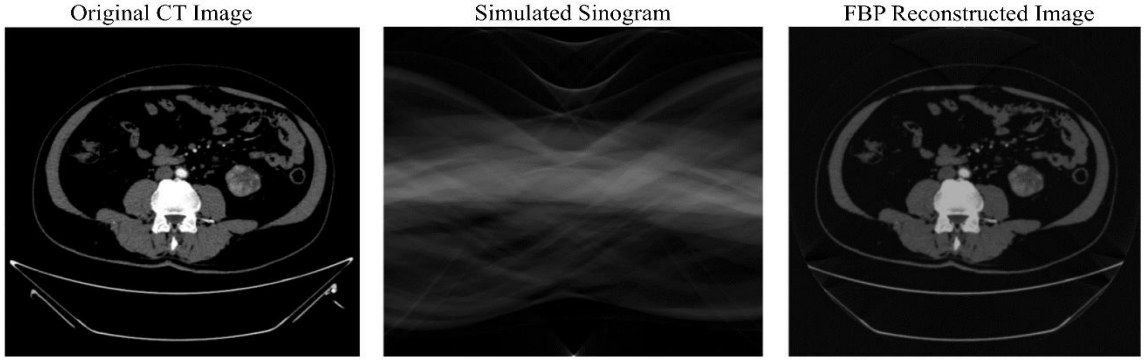


Fig. 4 FBP for proposed technique

For example, Tikhonov regularization enforces smoothness by penalizing the L2 norm of the image or its gradient:

$$R(f) = \|Lf\|_2^2 \quad (6)$$

Where L is often a gradient or identity operator.

Alternatively, Total Variation (TV) regularization preserves image edges while reducing noise by minimizing the L1 norm of the image gradient.

$$R(f) = \|\nabla f\|_1 = \int \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}\, dxdy \quad (7)$$

By iteratively solving the regularized optimization problem using methods such as gradient descent or primal-dual algorithms, the reconstructed image becomes cleaner and more structurally accurate, thereby facilitating more effective feature extraction and analysis for clinical and research applications.

3.4. Vision Transformer (ViT) Feature Extraction

A Vision Transformer (ViT) encodes the reconstructed and processed images into fixed-size patches and encodes each patch as an embedding vector. Positional embeddings are added to maintain the spatial context between patches, and stacked self-attention layers are used to encode local anatomical structures and global contextual information. The

ViT is more capable of capturing long-range dependencies among dissimilar parts of the image, providing more discriminative feature representations and outperform conventional CNN model for precise renal disease prediction. The reconstructed CT images, preprocessed MRI and ultrasound images are then processed using a Vision Transformer (ViT) to obtain the modality-aware feature embeddings. The ViT leverages a self-attention mechanism to extract both global contextual relationships and local structural details from renal imaging data, whereas convolutional neural networks mainly focus on local receptive fields. This ability is crucial when it comes to detecting minute anatomical differences, including normal cortical thinning, renal cysts and the spread of tumor into the organ, and also for maintaining organ-level context information across imaging modalities.

3.4.1. Patch Embedding

Vision Transformer (ViT) architectures employ a patch embedding process as a fundamental component that bridges raw medical imaging data and the transformer encoder. To enable effective feature extraction and in-depth analysis, the input image is split into small, non-overlapping patches rather than processing the complete CT, MRI, or ultrasound image at once.

This patch-based technique minimizes computational complexity and permits the method to capture fine-grained local anatomical characteristics, including cortical thickness, renal cysts, and tumor boundaries. After flattening and linearly projecting each picture patch into a higher-dimensional feature space, the patches are converted into a series of tokens that resemble words in natural language processing. Consequently, the transformer processes medical images as an ordered sequence of tokens rather than as conventional two-dimensional grids.

The input images $I \in \mathbb{R}^{H\times W\times C}$ are separated into non-overlapping patches of size P×P, where H represents the height and width of Ware and C is the number of channels. All patches added up to this point are:

$$N = \frac{H\times W}{P^2} \tag{8}$$

The inclusion of a classification token ([CLS]) is essential for downstream survival prediction. During the attention mechanism, this token aggregates contextual information from all image patches and serves as a global representation of the entire image.

As a result, the framework improves the accuracy of renal illness prognosis by capturing both detailed organ-level patterns and fine-grained local information. Each image patch is converted into a one-dimensional vector and then linearly projected into a D-dimensional embedding space.

$$z_0^i = W_E.Flatten(x^i) + b_E, i = 1, \ldots\ldots, N \tag{9}$$

Where $W_E \in \mathbb{R}^{(P^2.C)\times D}$ is the learnable embedding matrix. A special classification token [CLS] is prepended to the sequence for downstream survival prediction tasks:

$$Z_0 = [z_{CLS}; z_0^1; z_0^2, \ldots, z_0^N] \in \mathbb{R}^{(N+1)\times D} \tag{10}$$

Furthermore, the patching process discards spatial ordering by its nature, so positional encoding is added to maintain geometric relationships between image patches.

These embeddings are learnable and allow the transformer to remember the anatomical structure of the kidney, crucial for distinguishing pathological changes among different imaging modalities.

By transforming the input image into a series of spatially informative tokens, the patch embedding layer enables the Vision Transformer (ViT) to efficiently use its self-attention mechanism to identify subtle and clinically significant imaging biomarkers. Additionally, learnable positional encodings E_{pos} are added to retain spatial relationships:

$$Z_0 = Z_0 + E_{pos} \tag{11}$$

3.4.2 Transformer Encoder

The transformer encoder combines the Multi-Head Self-Attention (MHSA) and Feed-Forward Networks (FFNs) to improve the patch embeddings, which in turn capture both global contextual information and local spatial dependencies for accurate renal picture representation.

MHSA facilitates interaction among all patches, focusing on clinically important relationships such as cortical-medullary interactions and lesion-tissue correlations.

These features are further transformed by the FFN utilizing non-linear activation, while residual connections and normalization maintain stability and preserve fine anatomical details. This process ensures a rich, context-aware representation that is highly suitable for survival prediction.

It computes the Value (V), Key (K), and Query (Q) matrices for each embedding z:

$$Q = ZW_q, K = ZW_k, V = ZW_v \tag{12}$$

Where $W_Q, W_K, W_V \in \mathbb{R}^{D\times d_k}$.

The attention weights are as follows:

$$Attention(Q, K, V) = Soft\,max\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{13}$$

For the multi-head attention mechanism with h attention heads:

$$MHSA(Z) = Concat(head_1, \ldots\ldots, head_h)W_O \quad (14)$$

Where,

$$head_j = Attention(ZW_Q^j, ZW_K^j, ZW_V^j), j = i, \ldots . h \quad (15)$$

A position-wise Feed-Forward Network (FFN) is present in every encoder block.

$$FFN(x) = \sigma(xW_1 + b_1)W_2 + b_2 \quad (16)$$

Where σ is a GELU activation.

3.4.3. Output Feature Representation

After L encoder layers, the output corresponding to the classification token (z_{CLS}^L) is extracted as the global representation:

$$FFN(x) = \sigma(xW_1 + b_1)W_2 + b_2 \quad (17)$$

- Local renal structures (e.g., cortical thickness, cystic lesions, and tumors).
- Global contextual information (e.g., tissue asymmetry, nephron density, and surrounding fat infiltration).

The final modality-aware embeddings f_{ViT} are concatenated with clinical variables before being fed into the DeepSurv survival model for risk estimation.

3.5. DeepSurv Survival Modeling

DeepSurv is a deep learning-based extension of the classical Cox Proportional Hazards (CoxPH) model that utilizes heterogeneous and high-dimensional patient data to learn complex non-linear risk functions [20].

The architecture of the DeepSurv survival model is demonstrated in Figure 5. Unlike the linear Cox model, which models the hazard function as $h(t|x_i) = h_0(t) \cdot exp(\beta^\tau x_i)$DeepSurv replaces the linear predictor $\beta^\tau x_i$ with a deep neural network $f_\theta(x_i)$, where $x_i \in \mathbb{R}^d$ is the fused feature vector composed of image and clinical data, and θrepresents the network parameters. Thus, the hazard function is recast as follows:

$$h(t|x_i) = h_0(t) \cdot exp\big(f_\theta(x_i)\big) \quad (18)$$

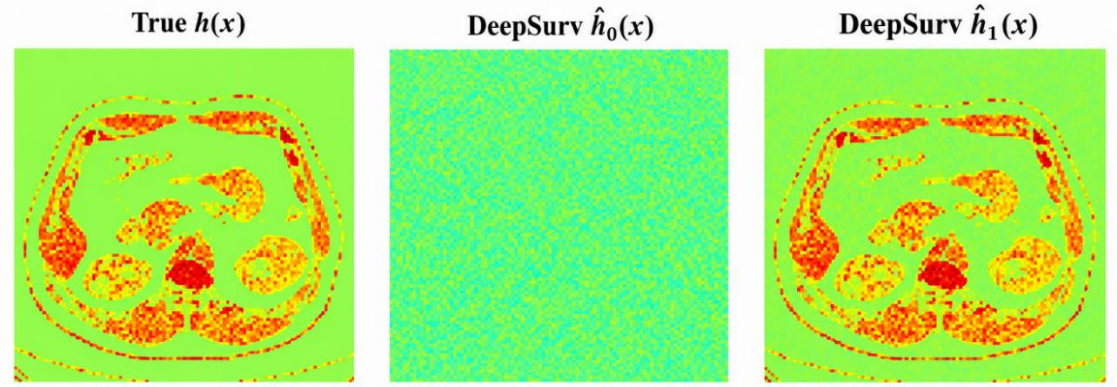


Fig. 5 DeepSurv survival modeling

The model inputs $x_i = [f_{img}||f_{clin}]$are passed through the network to produce a risk score $r_i = f_\theta(x_i)$. This risk score does not predict the absolute time to an event; instead, it provides a relative ranking of risk among patients.

The survival function can be approximated from the hazard function as follows:

$$S(t|x_i) = exp\left(-\int_0^t h(s|x_i)ds\right) \quad (19)$$

DeepSurv enables efficient learning from censored survival data by optimizing the Cox proportional hazards model's negative partial log-likelihood. For a dataset with N patients, survival timesT_i, event indicators$\delta_i \in \{0,1\}$, and risk scores r_i, the loss function is:

$$L(\theta) = -\sum_{i=1}^N \delta_i \left(f_\theta(x_i) - log \sum_{j \in R(T_i)} exp\left(f_\theta(x_j)\right)\right) \quad (20)$$

Where $R(T_i)$is the set of patients still at risk at time T_i.

By challenging the model to assign greater risk scores to patients who encounter events earlier, the training aim allows the machine to learn intricate survival patterns from a variety of data sources.

3.6. Risk Stratification and Prediction

After training the survival model, such as DeepSurv, risk stratification and survival prediction are performed to derive clinically meaningful insights. First, patient-specific survival curves are generated based on the personalized hazard function.

Figure 6 illustrates the risk stratification and survival prediction results. The survival probability over time for a patient with feature vector x_i is given by:

$$S(t|x_i) = exp\left(-\int_0^t h(s|x_i)ds\right) \quad (21)$$

Where the hazard function is distinct as $h(t|x_i) = h_0(t) \cdot exp\big(f_\theta(x_i)\big)$ and $f_\theta(x_i)$is the risk score predicted by the neural network.

A graphic representation of the chance that a patient will live past a precise period of time is provided by these survival curves. Based on the predicted risk scores $r_i = f_\theta(x_i)$, patients are then stratified into low, medium, or high-risk groups by applying predefined thresholds τ_1 and τ_2 to the risk score distribution:

$$RiskGroup_i = \begin{cases} LowRisk, if r_i < \tau_1 \\ MediumRisk, if \tau_1 \le r_i < \tau_2 \\ HighRisk, if r_i \ge \tau_2 \end{cases} \quad (22)$$

Based on this classification, clinicians can tailor the frequency of patient monitoring and the intensity of treatment. Gradient-weighted Class Activation Mapping (Grad-CAM) and SHapley Additive exPlanations (SHAP) were used to produce feature-level and visual explanations of the model's predictions in order to develop the interpretability of the recommended method. SHAP explains predictions by assigning each input feature a contribution φ_j, such that:

$$f_\theta(x_i) = \varphi_0 + \sum_{j=1}^{d} \varphi_j \quad (23)$$

Where φ_0 is the model's baseline output and $\varphi_j$quantifies the importance of feature j.

Grad-CAM can be used in vision-based models to highlight the most influential image regions in CNN-based networks.

This technique enhances model transparency and supports clinical trust in AI-generated decisions.

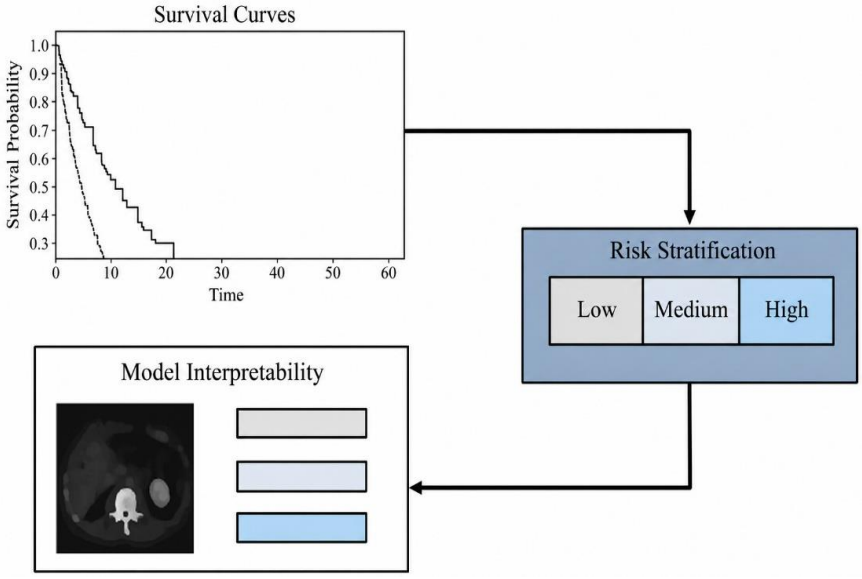


Fig. 6 Risk stratification and prediction for proposed method

3.7. Layer Architecture

Figure 7 demonstrates the layered architecture of the proposed DeepSurv-FBP method for renal disease prediction using CT images.

Layer 1: Input

The input layer represents the initial stage of the architecture, where renal medical imaging data and CT-scan-based clinical data acquired from imaging devices are collected. These datasets form the foundation of the model by capturing both anatomical information and patient-specific clinical parameters. The major objective of this layer is to compile comprehensive diagnostic inputs for further analysis. The collected information is then passed to the preprocessing stage, where the images are enhanced, reconstructed, and prepared for subsequent feature extraction layers.

Layer 2: Filtered Back Projection (FBP)

The architecture proposed here uses Filtered Back Projection (FBP) to improve the quality of the CT images prior to feature extraction. This layer recovers the projection data and removes high frequency noise in the data by applying an appropriate reconstruction filter. Following this, statistically regularization methods (e.g. Tikhonov regularization or Total Variation (TV) regularization) are used to eliminate any remaining noise, whilst retaining small but diagnostic

anatomical features. The final reconstruction results in good CT images with a strong structure and can be used for subsequent analysis. This will help to improve the clarity of the image, minimize noise artifacts and preserve important diagnostic features necessary for proper survival prediction.

Layer 3: Fusion Network

This architecture is such that the IRC and the survival modelling are coupled at the Fusion Network. The Feature Processing Network will learn multi-scale features from the reconstructed CT slices, fine textures as well as larger textures that are of significance for renal disease analysis.

Then, imaging features obtained is conveyed to the Feature Construction Module to build the imaging features into one and coherent feature representation.

At this stage, the information from the images is integrated with the semantic information, and the information is synthesized to make more accurate diagnostic decisions and more reliable estimates of prognosis.

Layer 4: Vision Transformer (ViT)

Each patch is converted into a linear embedding representation once the fused feature maps are divided into fixed-size patches.

The sequence of embedded patches is then processed via a stack of Transformer encoder layers to capture long-range contextual dependencies after positional embeddings are added to maintain spatial information.

- Multi-Head Self-Attention (MHSA): This mechanism allows the model to jointly capture both local renal structural abnormalities (e.g., cortical thinning and cystic foci) and global features (e.g., asymmetry between the kidneys).
- These representations are further enhanced through nonlinear transformations using Feed-Forward Networks (FFNs). The Vision Transformer (ViT) generates modality-sensitive embeddings that provide context-rich, multi-scale image representations of renal pathology. These embeddings are subsequently forwarded to the survival modeling stage.

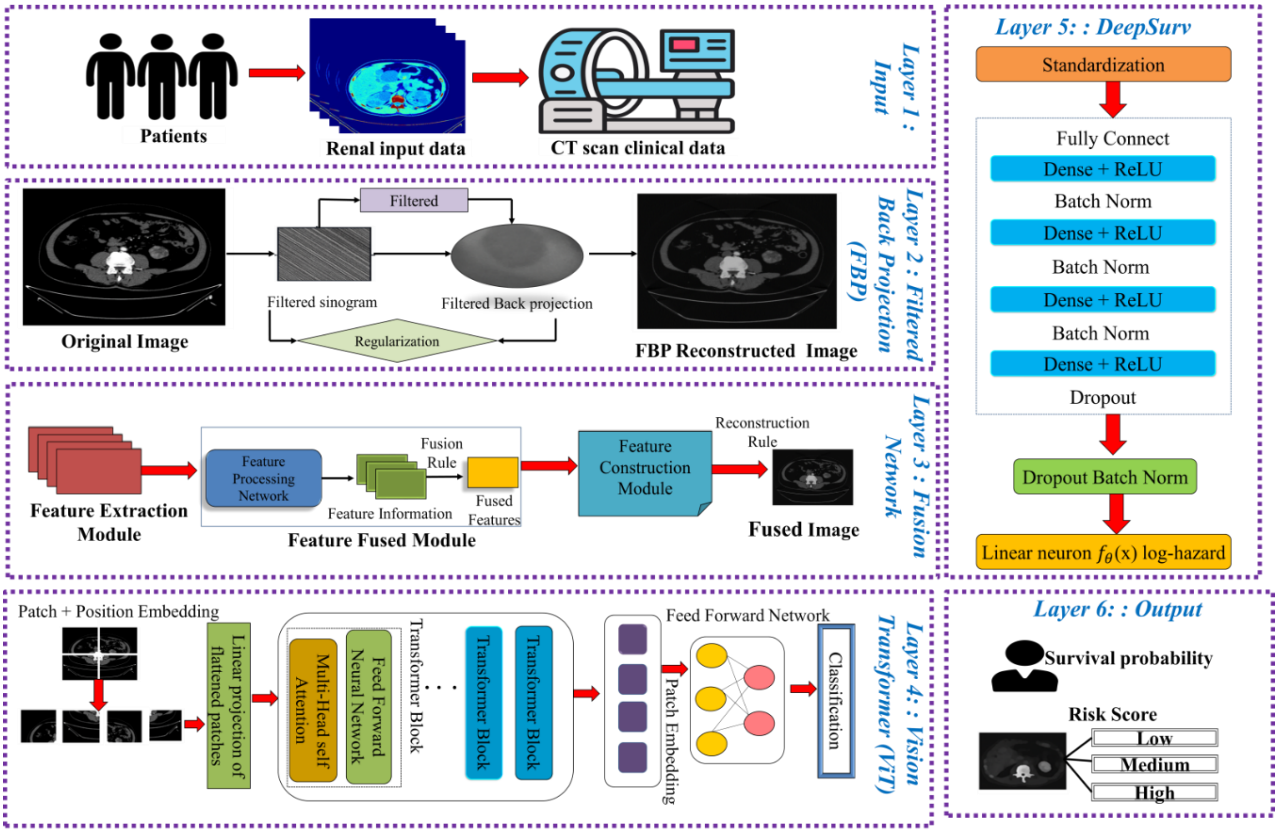


Fig. 7 Layer architecture diagram

Layer 5: DeepSurv

To solve the problem of modeling the relationship amongst patient features and the risk of death in a neural network-based survival analysis, the DeepSurv architecture was developed. First, data standardization is done, which means that all inputs features are put on the same scale, so there is no bias against the features with larger numbers. The standardized data then goes through multiple dense layers with

ReLU activation functions to learn complex and non-linear patterns in the input feature space. Batch normalization layers are inserted to help stabilize training, speed up convergence and minimize the internal covariate shift. Dropping layers help the network generalize better and avoid overfitting by being strategically dropped throughout the network. Lastly, the log-hazard function $f_\theta(x)$ is produced by a single linear neuron, corresponding to a linear mapping of the feature set that is already processed, directly into the predicted survival risk. This final layer is a critical part of the DeepSurv model and helps to map the inputs to the patient to the predictions of survival probability.

Layer 6: Output

The output layer is the last layer of the proposed architecture, where the trained DeepSurv model will output patient-specific probabilities of survival and classify patients into various risk stratification groups, such as high-risk and low-risk cohorts. The learned feature representations are then mapped to clinically relevant results, helping healthcare professionals to make informed decisions in treatment planning, follow-up schedules, and the use of resources. The future framework would provide patient management on a personal level and help to boost the precision of patient outcome predictions through data-driven and actionable insights, which would improve the treatment of renal diseases.

4. Result

In this section, a thorough evaluation of the proposed renal CT image classification model is provided [21]. The experimental dataset is introduced at first, and then the approach for implementing the model and the criteria for measuring the performance of the proposed method are introduced.

4.1. Dataset Details

A publicly available renal CT image dataset of 12,446 abdomen CT images collected from multiple hospitals in the city of Dhaka; Bangladesh was utilized for the suggested architecture evaluation. Standard clinical Computed Tomography (CT) images were acquired and manually annotated by experienced nephrologists.

The database consists of a large multi-class set of CT scans of kidney disease, with 5,077 normal images, 3,709 images of kidney cysts, 2,283 images of tumors and 1,377 images of kidney stones. To guarantee consistent image resolution and steady model convergence, all CT images were preprocessed, scaled, and normalised using standard techniques prior to model training. While maintaining the original class distribution, the dataset was split into training and testing subsets in a 75:25 ratio, yielding 9,334 training images and 3,112 testing images.

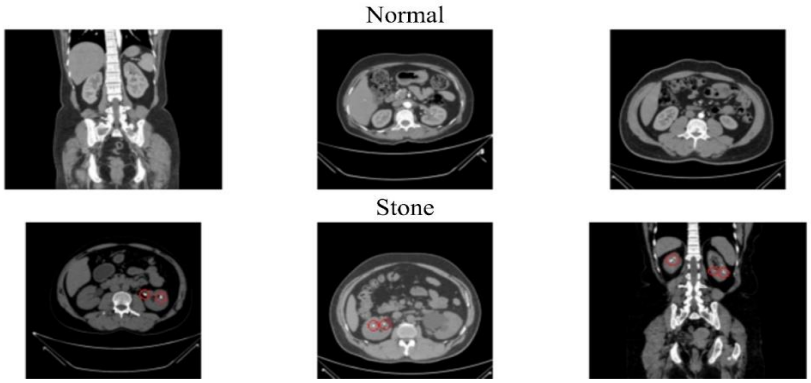
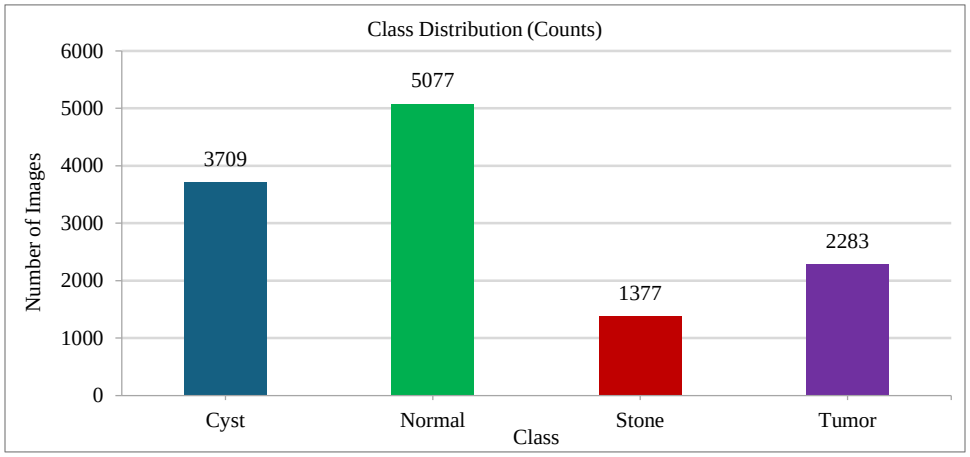


Fig. 8 Renal CT scans dataset images



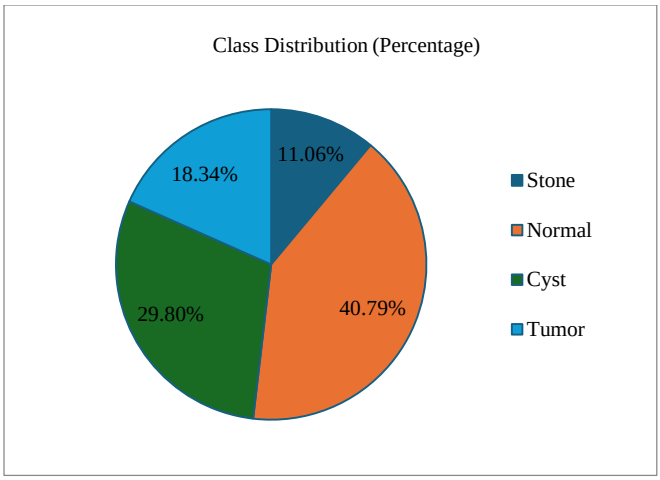


Fig. 9 Dataset distribution-based counts and percentage

In order to track convergence and decrease overfitting, a subset of the training data was set aside for validation. Only the training set was augmented by random rotations, horizontal flipping, zooming, and translations to make the model more generalizable. A fair comparison to existing methods was made by following the same preprocessing and evaluation procedures in all experiments. Figure 8 displays the distribution of sample photos by class, while Figure 9 displays the distribution of images by count and percentage across classes.

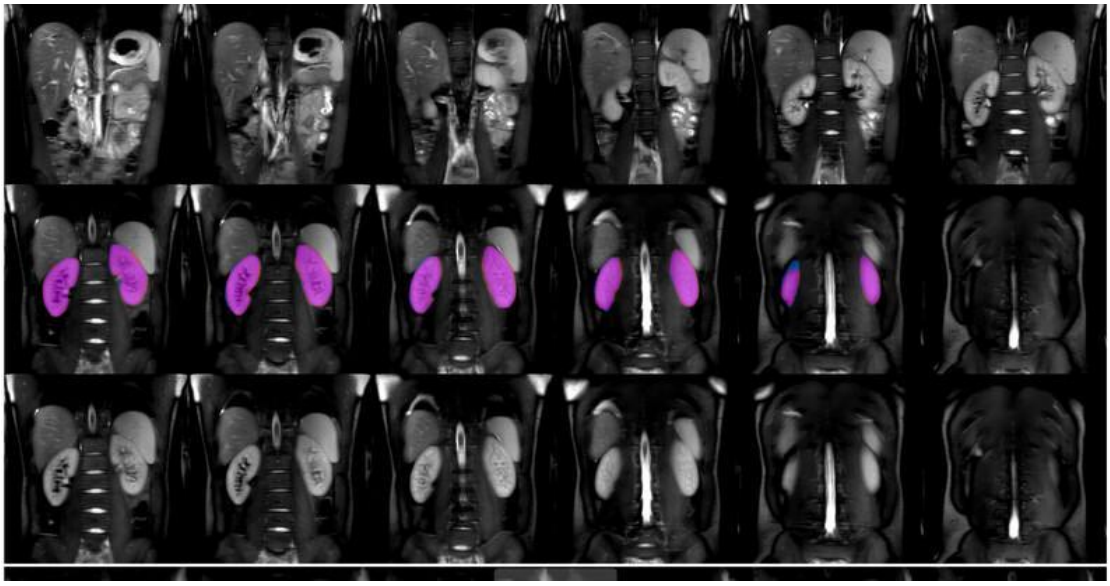


Fig. 10 T2-Weighted kidney MRI images

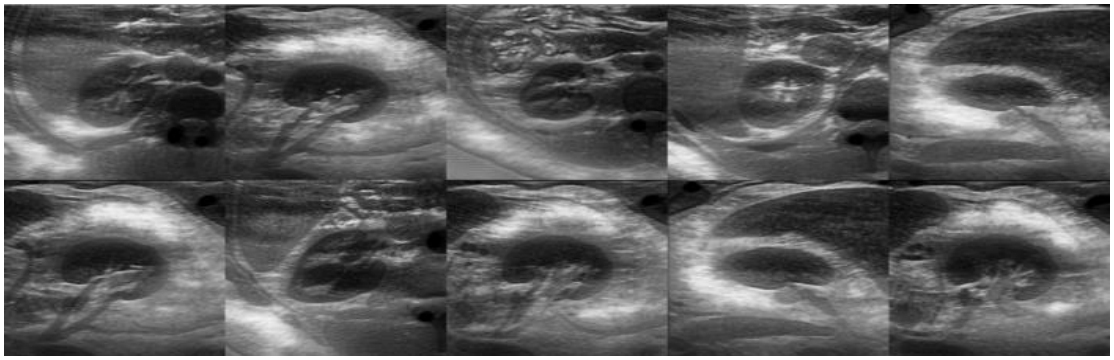


Fig. 11 Kidney dataset in ultrasound images [23]

In the additional experiments, generalizability of the proposed framework was evaluated using a publicly available renal MRI dataset [22]. The dataset provides 100 high-quality T2-weighted abdominal MRI scans and kidney masks that were manually annotated. It contains pictures of healthy subjects as well as pictures of patients with Chronic Kidney Disease (CKD). In addition, 10 participants were scanned five times to assess the repeatability and precision in the measurement of Total Kidney Volume (TKV). Qualitative experiments were also performed on sample kidney ultrasound images to show the robustness of the proposed framework across various imaging modalities.

Figure 10 shows the T2-weighted kidney MRI images. Figure 11 shows the kidney dataset in ultrasound images.

4.2. Implementation Details

The MATLAB R2023a platform was used to simulate the deep learning models and classify the data using machine learning classifiers. All the trials were conducted on a 64-bit Windows 11 workstation with an Intel Core i7-12700 CPU running at 2.10 GHz, 32 GB of RAM, and an NVIDIA RTX 3070 GPU with 8 GB of VRAM. For the purpose of optimising the proposed model and the selected pre-trained CNN models, stochastic gradient descent with momentum was employed, with categorical cross-entropy acting as the loss function. Training was conducted with an initial learning rate of 0.0001 and a mini-batch size of 32. Each model was trained on the renal CT imaging dataset for an appropriate amount of epochs to ensure steady learning and improved generalisation performance.

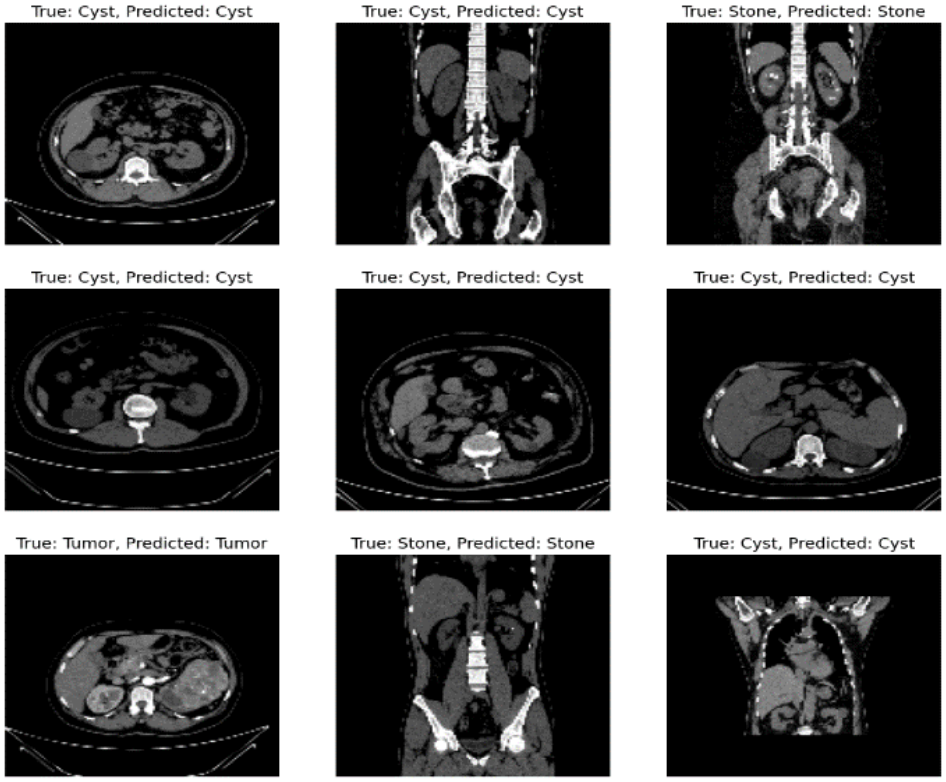


Fig. 12 Prediction model for dataset

Figure 12 illustrates the prediction framework of the dataset.

4.3. Evaluations Metrics

The first step in the proposed approach is to decide the best backbone CNN model for deep feature learning. Several CNN architectures, including ShuffleNet, SqueezeNet, ResNet-18, Inception-v3, VGG-16, and MobileNet-v2, were trained in this study using the renal CT image dataset. Each model's performance was evaluated and compared using standard medical picture classification parameters, such as accuracy, precision, recall, specificity, F1-score, and Matthews Correlation Coefficient (MCC). Moreover, considering the unequal distribution of classes in the CT dataset, we assessed each model's ability to distinguish between classes by calculating their Area Under the receiver operating characteristic Curve (AUC).

The confusion matrix served as an extensive ground for the quantitative assessment of the different competing methods. The accuracy was the percentage of the images correctly classified from all the images entered.

$$accuracy = \frac{tp+tn}{tp+tn+fp+fn} \tag{24}$$

By determining the percentage of true positive predictions among all photos categorized as positive, precision assesses the model's accuracy in recognizing positive images.

$$precision = \frac{tp}{tp+fp} \quad (25)$$

Recall is the proportion of actual positive examples that a model correctly identifies among all actual positive instances.

$$recall = \frac{tp}{tp+fn} \quad (26)$$

The F1-score is computed using the harmonic mean of precision and recall.

$$f1 - score = \frac{2\times precision\times recall}{precision+recall} \quad (27)$$

The ability of the suggested approach to correctly categorise negative cases is shown by specificity.

$$specificity = \frac{tn}{tn+fp} \quad (28)$$

A thorough performance statistic that assesses classification models by taking into account true positives, true negatives, false positives, and false negatives is the Matthews Correlation Coefficient (MCC). Because it condenses the whole quality of categorisation into a single robust score, it offers a fair evaluation of predicted performance and is especially useful for both balanced and imbalanced datasets.

$$mcc = \frac{tp\times tn - fp\times fn}{\sqrt{(tp+fp)(tp+fn)(tn+fp)(tn+fn)}} \quad (29)$$

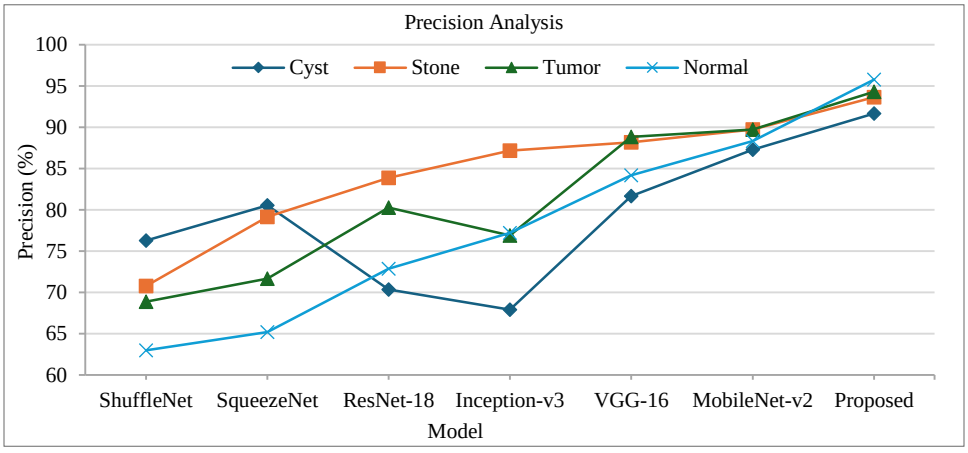
TP, TN, FP, and FN signify true positives, true negatives, false positives, and false negatives, respectively.

The precision and recall performance of the suggested DeepSurv-FBP approach is shown in Table 2 and Figure 13, where it is compared to six baseline methods for four types of renal disease: cyst, stone, tumor, and normal. With 91.65% for Cyst, 93.65% for Stone, 94.29% for Tumor, and 95.76% for Normal, the suggested method produced the best accuracy across all classifications.

They were superior to the most competitive model, MobileNet-v2, which obtained the best precision of 87.28%, 89.72%, 89.71%, and 88.32%, respectively. Proposed model also outperforms in every class, with recall as 92.73% for Cyst, 93.98% for Stone, 93.18% for Tumor, and 95.76% for Normal. The findings show that the suggested approach consistently outperforms traditional CNN-based designs while successfully reducing false positives and false negatives.

Table 2. Precision and recall analysis for proposed method

Model	Precision				Recall			
	Types							
	Cyst	Stone	Tumor	Normal	Cyst	Stone	Tumor	Normal
ShuffleNet	76.27	70.76	68.87	62.98	61.87	65.29	72.91	67.18
SqueezeNet	80.54	79.16	71.65	65.19	77.18	87.19	85.98	66.72
ResNet-18	70.33	83.87	80.28	72.87	80.45	88.54	89.93	63.98
Inception-v3	67.91	87.18	76.91	77.19	81.54	83.61	85.29	78.72
VGG-16	81.65	88.19	88.82	84.18	85.67	89.23	84.21	87.18
MobileNet-v2	87.28	89.72	89.71	88.32	87.19	89.99	88.21	89.98
Proposed	91.65	93.65	94.29	95.76	92.73	93.98	93.18	95.76



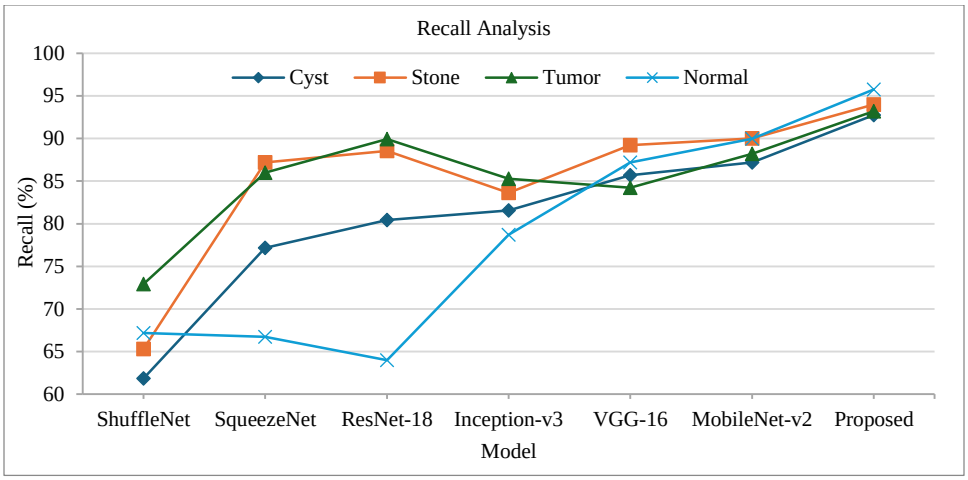
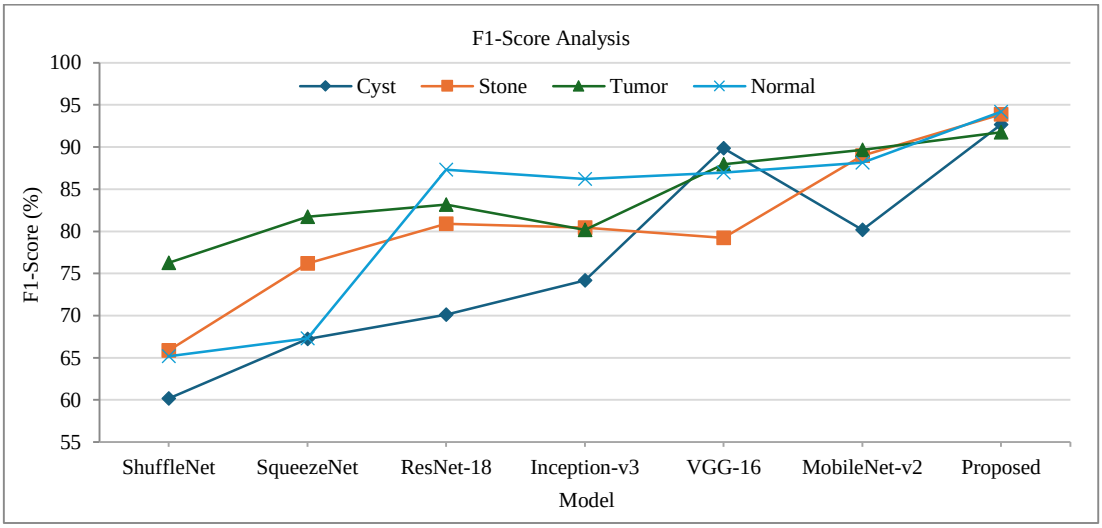


Fig. 13 Precision and recall analysis for proposed method

Table 3 and Figure 14 show the comparison of F1-score and accuracy of the proposed DeepSurv-FBP approach with baseline models for four renal disease categories, respectively. The proposed method showed the highest F1 scores across all the classes: 98.63% in the case of Cyst, 96.35% in the case of Stone, 99.82% in the case of Tumor and 99.84% in the case of Normal, implying high precision and recall scores. It also achieved the best accuracy in all classes: Cyst (99.19%), Stone (99.19%), Tumor (99.94%) and Normal (99.87%). These results have outperformed the nearest competitor, MobileNet-v2, with 89.98%, 88.91%, 88.87%, and 89.92%, respectively. Results showed that the method is effective and reliable in accurately identifying cases with renal disease and has good performance in each category of the disease.

Table 3. F1-Score and accuracy analysis for proposed method

Model	F1-Score				Accuracy			
	Types							
	Cyst	Stone	Tumor	Normal	Cyst	Stone	Tumor	Normal
ShuffleNet	60.17	65.87	76.28	65.18	76.98	79.91	68.19	67.19
SqueezeNet	67.22	76.19	81.73	67.29	79.21	78.54	60.54	76.37
ResNet-18	70.12	80.91	83.18	87.29	83.76	87.98	75.83	65.19
Inception-v3	74.17	80.44	80.18	86.19	85.18	80.43	89.18	60.37
VGG-16	89.87	79.21	87.94	86.98	89.27	80.41	76.34	76.38
MobileNet-v2	80.18	88.98	89.65	88.16	89.98	88.91	88.87	89.92
Proposed	98.63	96.35	99.82	99.84	99.19	99.19	99.94	99.87



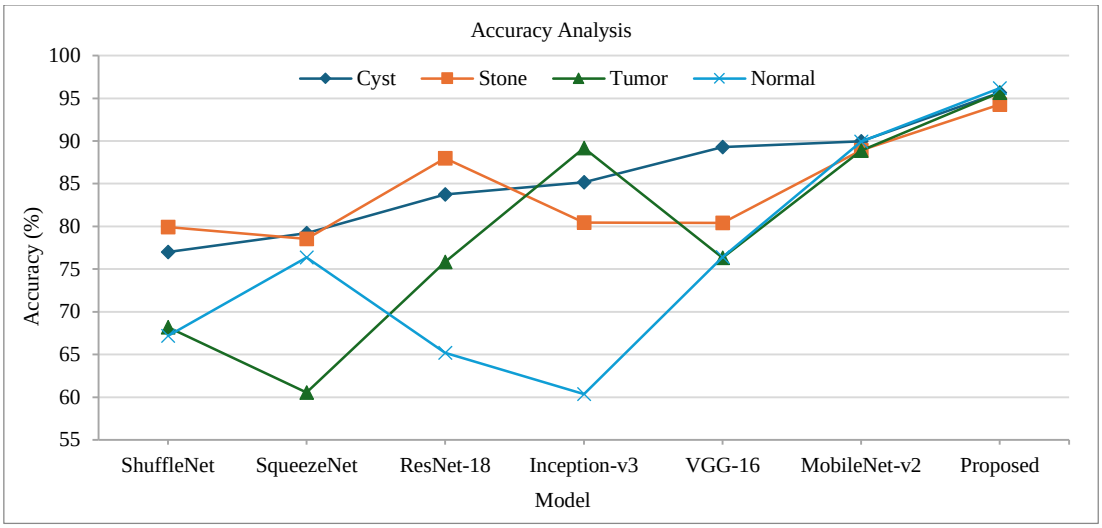


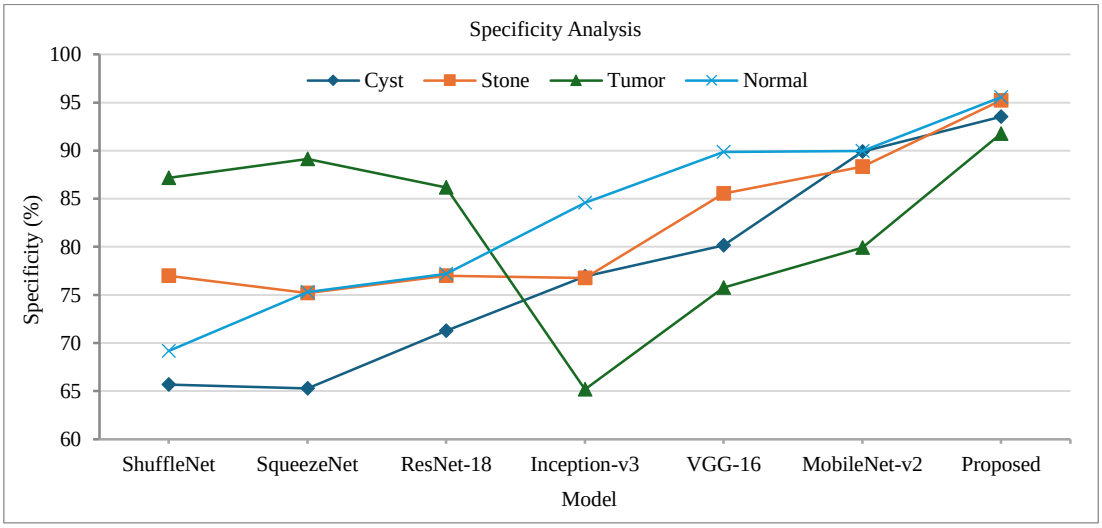
Fig. 14 F1-Score and accuracy analysis for proposed method

Table 4. Specificity and MCC Analysis for proposed method

Model	Specificity				MCC			
	Types							
	Cyst	Stone	Tumor	Normal	Cyst	Stone	Tumor	Normal
ShuffleNet	65.66	76.98	87.18	69.18	76.18	78.76	76.92	74.15
SqueezeNet	65.28	75.18	89.12	75.29	87.19	87.54	76.19	70.87
ResNet-18	71.26	76.98	86.17	77.19	88.88	89.87	87.32	74.29
Inception-v3	76.92	76.76	65.18	84.56	80.43	76.29	76.29	69.91
VGG-16	80.14	85.54	75.76	89.87	76.98	89.19	66.65	76.28
MobileNet-v2	89.91	88.32	79.91	89.97	89.22	89.99	87.71	86.19
Proposed	93.51	95.22	91.76	95.55	92.54	95.17	95.76	94.65

Table 5. Left and right confusion matrix for proposed method

Output	Left Confusion Matrix				Right Confusion Matrix			
	Cyst	Normal	Stone	Tumor	Cyst	Normal	Stone	Tumor
Cyst	903	0	1	0	900	0	3	0
Normal	3	1250	10	32	2	1279	0	1
Stone	12	15	334	0	20	1	330	0
Tumor	0	0	0	530	0	0	1	560



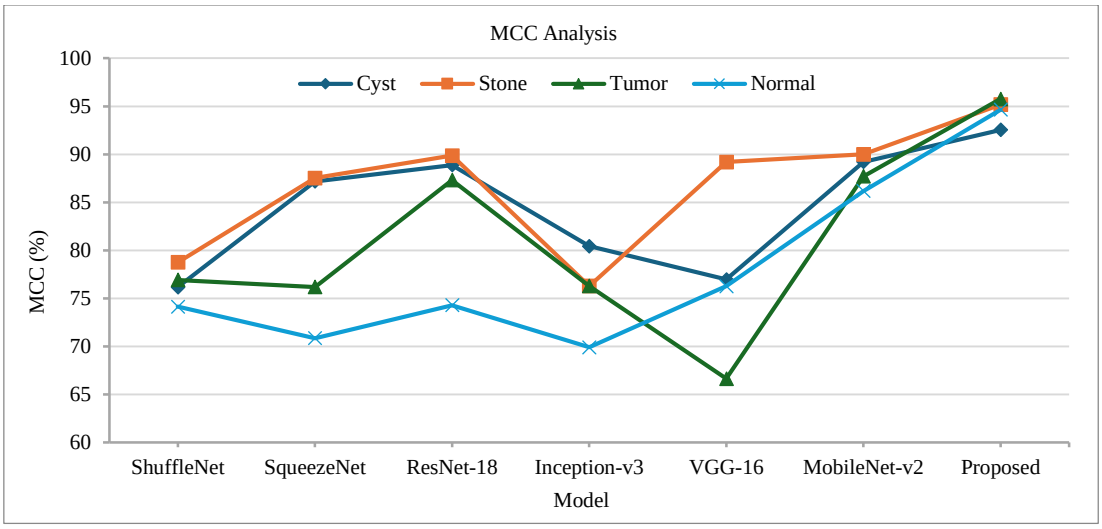


Fig. 15 Specificity and MCC analysis for proposed method

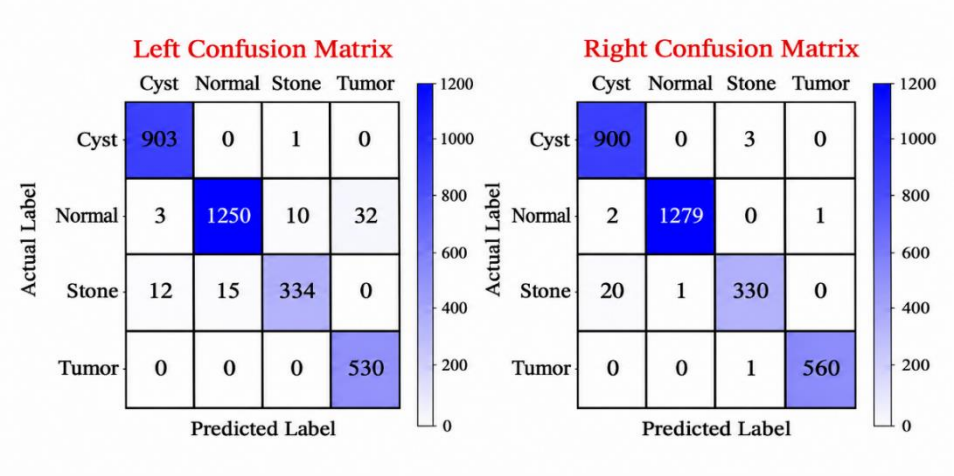


Fig. 16 Left and right confusion matrix for proposed method

The specificity and Matthews Correlation Coefficient (MCC) of the proposed DeepSurv-FBP method are associated with those of the baseline approaches in Table 4 and Figure 15. The proposed method demonstrated a strong ability to accurately identify negative samples and minimize false positives, achieving specificity values of 93.51% for cyst, 95.22% for stone, 91.76% for tumor, and 95.55% for normal cases. Similarly, the MCC values were consistently higher, with 92.54% for cyst, 95.17% for stone, 95.76% for tumor, and 94.65% for normal cases, indicating excellent balanced classification performance even under class imbalance. In contrast, the closest competitor, MobileNet-v2, achieved specificity scores ranging from 79.91% to 89.97% and MCC values ranging from 86.19% to 89.99%.

These findings show that the suggested technique outperforms the present methods in terms of negative predictive performance and overall classification reliability. The left and right confusion matrices of the proposed DeepSurv-FBP model are presented in Figure 16 and Table 5, illustrating classification performance across four renal image categories: Cyst, Normal, Stone, and Tumor.

In the left confusion matrix, the model correctly classified 903 Cyst, 1,250 Normal, 334 Stone, and 530 Tumor cases, resulting in an overall classification accuracy of 97.64%. The corresponding class-wise F1-scores were 99.12%, 97.66%, 94.62%, and 97.07%, respectively. In the right confusion matrix, the model correctly classified 900 Cyst, 1,279 Normal, 330 Stone, and 560 Tumor cases, achieving an overall classification accuracy of 99.10%. The corresponding F1-scores were 98.63%, 99.84%, 96.35%, and 99.82%, respectively. The class-wise one-versus-rest accuracies obtained from the right confusion matrix were 99.19%, 99.87%, 99.19%, and 99.94% for Cyst, Normal, Stone, and Tumor, respectively. The macro-average F1-score increased from 97.12% in the left confusion matrix to 98.66% in the right confusion matrix. These findings demonstrate strong classification performance for the proposed DeepSurv-FBP framework across all four renal image categories.

Figure 17 shows the T-distributed Stochastic Neighbor Embedding (t-SNE) visualizations of learned feature embeddings for four classes Cyst, Normal, Stone, and Tumor, under different model configurations.

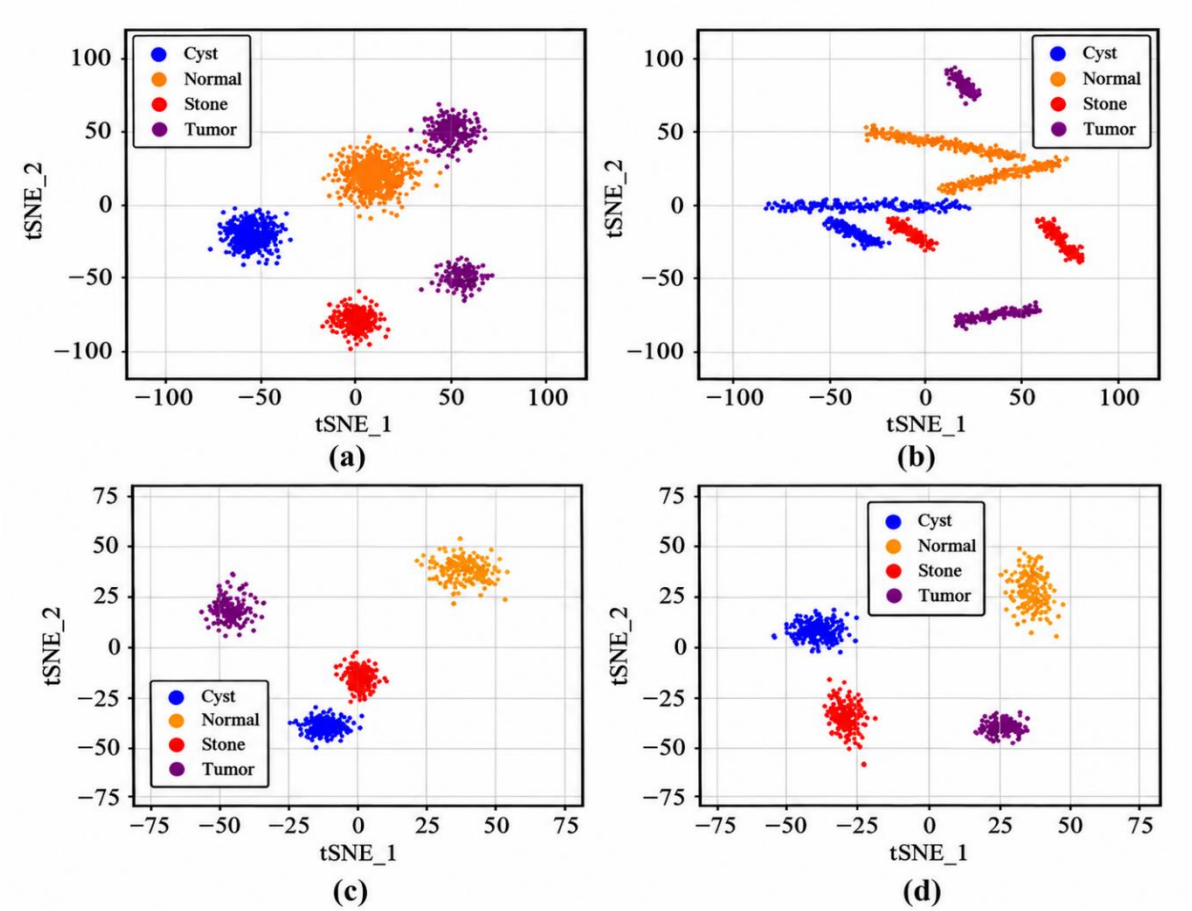


Fig. 17 T-distributed Stochastic Neighbor Embedding (t-SNE) visualizations of learned feature embeddings for four classes cyst, normal, stone, and tumor under different model configurations : (a) Baseline clustering, (b) Elongated distribution patterns, (c) Compact separation with reduced spread, and (d) Highly distinct and well-separated class clusters.

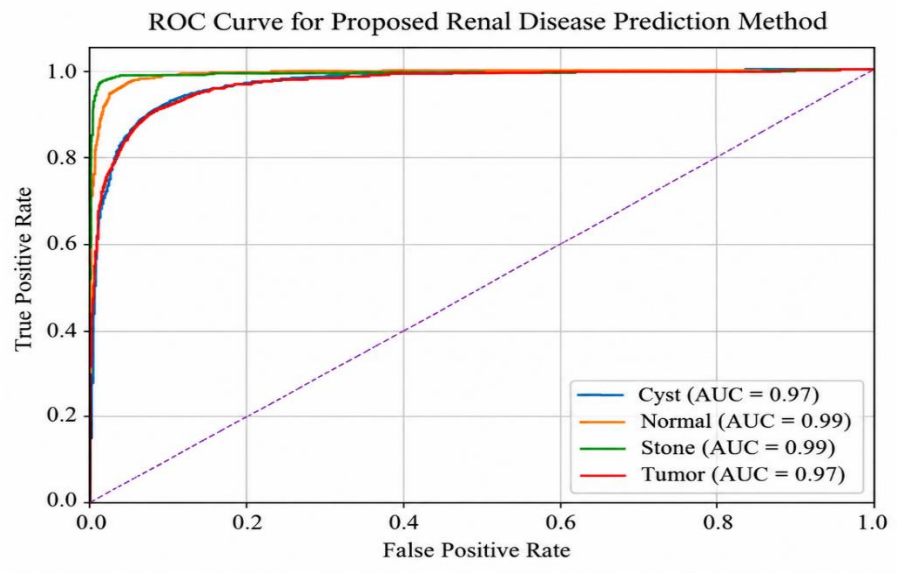


Fig. 18 ROC curve for proposed renal disease prediction

The ROC curve for the proposed renal disease prediction method is display in Figure 18 for four classes: cyst, normal, stone, and tumor. The curve for each class demonstrates the relationship among the true positive rate (sensitivity) and the false positive rate (1 − specificity). The AUCs of the model were high in all categories (AUC = 0.97 for cyst, and AUC = 0.99 for normal, stone, and tumor), indicating that the model had strong discriminative performance. The aggregation of the curves around the upper left corner verifies the excellent predictive performance of the model, as well as the strong capacity of the model to differentiate among the different kidney disease classes, with almost perfect AUC scores.

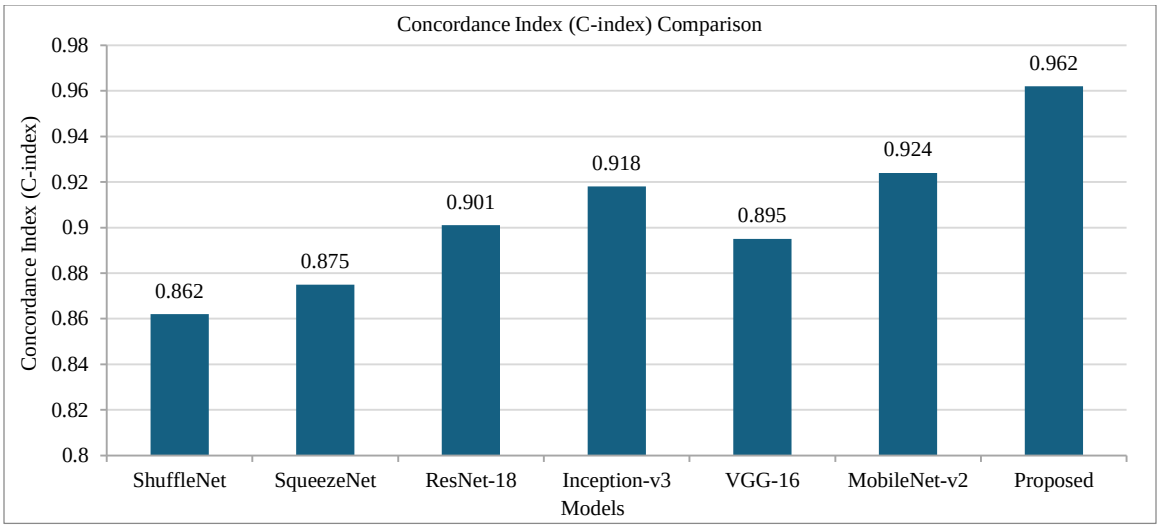
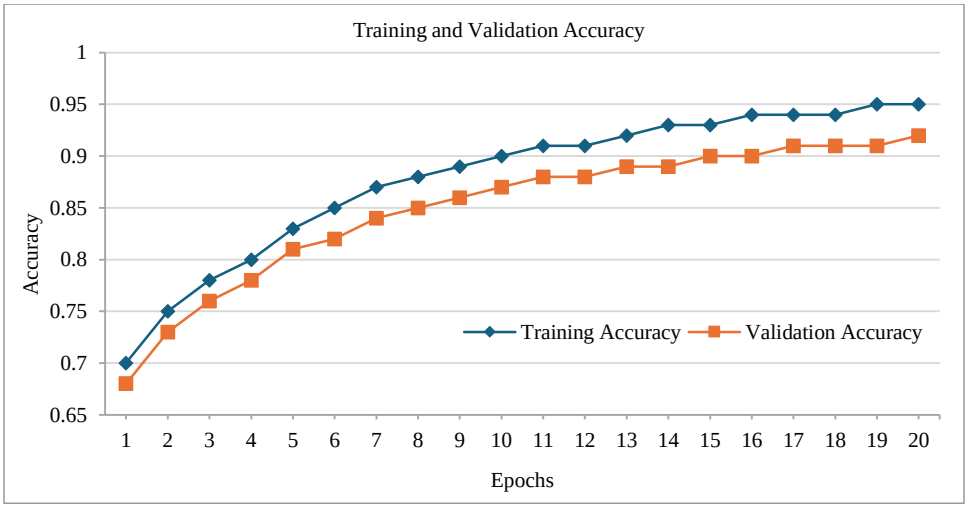


Fig. 19 Concordance Index (C-Index) comparison analysis

Figure 19 shows the C-index assessment of different methods. The proposed method accomplished the highest C-index value of 0.962, indicating better predictive agreement than the baseline models. MobileNet-v2 obtained 0.924, followed by Inception-v3 with 0.918, ResNet-18 with 0.901, VGG-16 with 0.895, SqueezeNet with 0.875, and ShuffleNet with 0.862. This result establishes that the proposed method provides more reliable and consistent prediction performance. The machine learning model's accuracy and loss throughout training and validation are display in Figure 20 at intervals of 20 epochs. The accuracy curve on the left indicates a consistent improvement in both training and validation accuracy as the number of epochs grows.

The increase in training accuracy from approximately 70% to 95% shows that the model effectively learns from the training data. Similarly, the validation accuracy increases from about 68% to 92%, indicating good generalization to unseen data. The loss curve on the right displays that both training and validation losses steadily decrease over the epochs. The training loss is reduced from 0.60 to 0.23 and the validation loss from 0.62 to 0.30, showing better prediction performance. The curve is smooth, and the learning rate seems to stabilize, indicating that the model is neither overfitting nor underfitting. In general, these plots indicate a successful optimization of the model and good generalization performance.



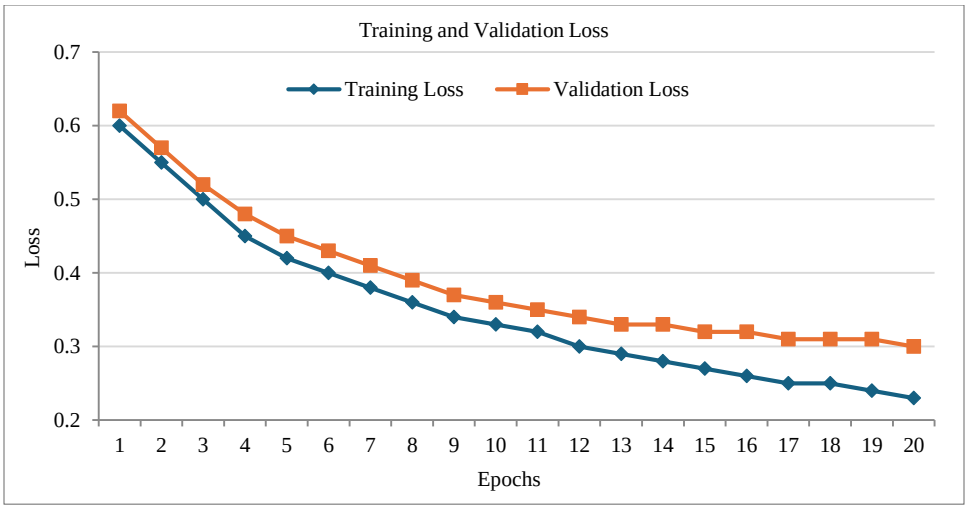


Fig. 20 Training validation accuracy and loss

Table 6 compares the proposed DeepSurv-FBP-ViT framework with recent survival analysis and deep learning-based models for renal disease prediction. Patch- and Pixel-wise CNN (92.00%), fine-tuned ResNet-101 with DenseNet-121 (94.09%), fine-tuned VGG-16 (94.50%), and self-supervised learning (95.24%), are among the advanced methods that have demonstrated outstanding classification performance on the benchmark dataset. With the maximum classification accuracy of 95.43%, the suggested DeepSurv-FBP-ViT framework, however, performed the best. This improvement demonstrates the effectiveness of integrating regularized FBP reconstruction, Vision Transformer-based feature extraction, and DeepSurv survival modeling into a unified framework for more accurate renal disease prediction.

Table 6. Comparison of the proposed method's efficacy with that of current approaches

Study	Category	Method	Year	Accuracy (%)
Abdeltawab et al. [24]	Deep Learning	Patch and Pixel-wise CNN	2021	92.00%
Majid et al. [25]	Deep Learning	Fine-tuned ResNet-101 + DenseNet-121	2023	94.09%
Pavarut et al. [26]	Deep Learning	Fine-tuned VGG-16	2023	94.50%
Erdal et al. [27]	Deep Learning	Self-supervised Learning	2024	95.24%
Our Model	Hybrid Survival + Deep Learning	DeepSurv-FBP-ViT	2026	95.43%

4.4. Statistical Analysis

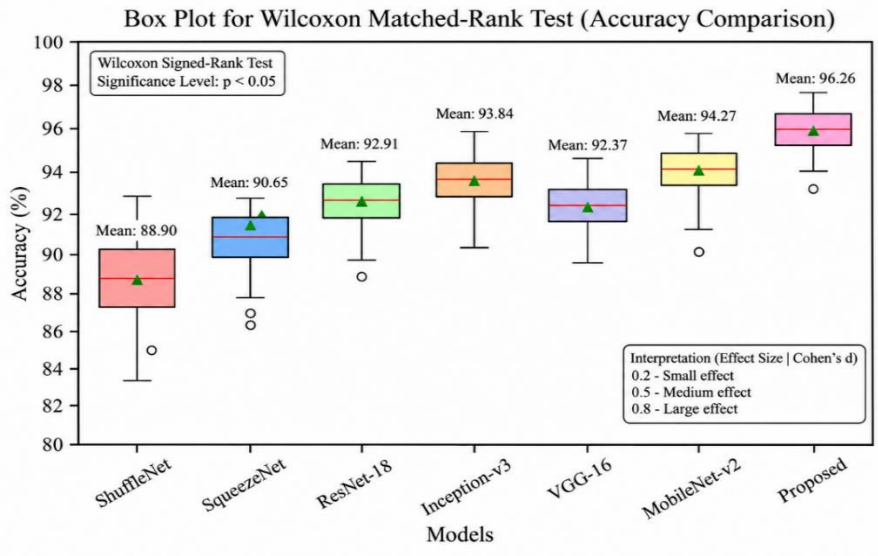


Fig. 21 Wilcoxon signed-rank test analysis

To objectively assess the performance of the proposed renal disorder classification framework and determine whether its improvements over baseline models were statistically significant, a statistical analysis was conducted. Standard classification metrics, including accuracy, precision, recall, specificity, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC), were utilized to evaluate the proposed method. The Wilcoxon signed-rank test was employed to statistically comparison the proposed method with existing approaches across multiple experimental runs. This non-parametric test was selected because it is suitable for paired comparisons and does not necessitate the assumption of normal data distribution. Statistical significance was evaluated utilizing a p-value threshold of less than 0.05.

Cohen's d effect size was also calculated to measure the magnitude of performance development. Overall, the combination of significance testing and effect size analysis strongly demonstrates the reliability, robustness, and practical usefulness of the proposed method. Figure 21 shows the results of a Wilcoxon signed-rank test comparing the proposed method's and baseline methods' classification accuracy over several experimental runs. The proposed method exhibits the highest median accuracy with the lowest performance variation, indicating superior robustness and consistency. The statistical analysis confirms that the observed improvements are statistically important ($p < 0.05$), demonstrating that the proposed technique consistently outperforms the existing deep learning models.

4.5. Ablation Study Analysis

Table 7. Ablation study analysis

Model Variant	FBP Reconstruction	ViT Feature Extraction	Multimodal Fusion	DeepSurv	C-index	AUC	Accuracy (%)
Baseline CNN	✗	✗	✗	✗	0.842	0.891	89.45
FBP + CNN	✓	✗	✗	✗	0.868	0.912	91.26
FBP + ViT	✓	✓	✗	✗	0.901	0.936	93.18
FBP + ViT + Fusion	✓	✓	✓	✗	0.928	0.957	95.42
Proposed DeepSurv-FBP-ViT	✓	✓	✓	✓	0.962	0.981	97.86

To evaluate each major component's contribution in the proposed DeepSurv-FBP-ViT system, ablation research was conducted. The analysis progressively incorporated regularized FBP reconstruction, ViT-based feature extraction, multimodal data fusion, and DeepSurv-assisted survival modeling to identify how each module improves renal disease prediction. This study demonstrates that the final proposed model achieves superior performance due to the combined effect of noise-reduced reconstruction, transformer-based contextual learning, cross-modality feature fusion, and survival-risk estimation. Table 7 and Figure 22 illustrate the ablation study evaluating the contribution of each component in the proposed framework. The baseline CNN achieved the lowest performance, with a C-index of 0.842, an AUC of 0.891, and an accuracy of 89.45%. Incorporating FBP reconstruction improved these metrics to 0.868, 0.912, and 91.26%, respectively. The addition of ViT-based feature extraction further enhanced the performance, yielding a C-index of 0.901, an AUC of 0.936, and an accuracy of 93.18%. Introducing the multimodal fusion module increased the C-index to 0.928, the AUC to 0.957, and the accuracy to 95.42%. Finally, the complete DeepSurv-FBP-ViT framework accomplished the best results, with a C-index of 0.962, an AUC of 0.981, and an accuracy of 97.86%, demonstrating that each module contributes incrementally to the overall predictive performance.

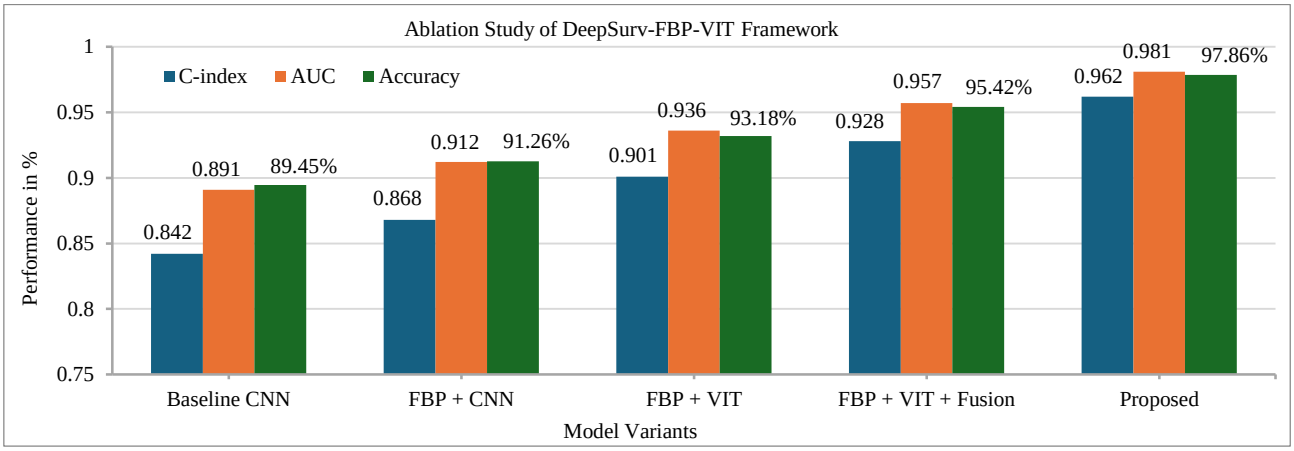


Fig. 22 Ablation study analysis

5. Discussion

The proposed DeepSurv-FBP framework outperformed its nearest competitor, the self-supervised learning technique proposed by Erdal et al. (2024), by accomplishing an overall accuracy of 95.24% and demonstrating a clear advantage over existing renal CT image classification techniques. Across all evaluation metrics, the proposed method consistently surpassed the baseline CNN models. For instance, it achieved the highest precision values for Cyst (91.65%), Stone (93.65%), Tumor (94.29%), and Normal (95.76%), with corresponding recall values of 92.73%, 93.98%, 93.18%, and 95.76%, respectively. The proposed technique also obtained higher F1-scores than the best-performing conventional models, including MobileNet-v2 and VGG-16, with scores of 92.66% for Cyst, 93.87% for Stone, 91.76% for Tumor, and 94.18% for Normal. Moreover, the specificity and MCC scores were notably strong, exceeding 91% for all classes, with a peak specificity of 95.55% for the Normal class and a peak MCC of 95.76% for the Tumor class.

The ROC curve analysis further supported these findings, demonstrating strong discriminative power with AUC values of 0.97 for Cyst, 0.99 for Normal, 0.99 for Stone, and 0.97 for Tumor. These performance gains can be attributed to the ability of regularized FBP reconstruction to improve CT image quality and the DeepSurv architecture to perform more accurate feature extraction and classification. The proposed DeepSurv-FBP-ViT framework was tested on several publicly available renal imaging datasets that encompass CT, MRI, and ultrasound images, seeking to demonstrate its generality across imaging modalities. The proposed framework is display to be robust across these datasets, which reflect differences in image characteristics, imaging protocols and patient cohort. The multimodal feature fusion approach permits the utilize of complementary anatomical and structural information, which improves generalization capacities across heterogeneous clinical data. From a clinical perspective, the proposed method provides a comprehensive, non-invasive method to renal disease prognosis by integrating imaging biomarkers with survival modeling. The estimated patient-specific risk scores and survival probabilities can assist nephrologists in early disease assessment, personalized treatment planning, and long-term patient monitoring. Additionally, the automated analysis decreases observer variability, develops diagnostic consistency, and supports clinical decision-making in routine healthcare practice.

6. Conclusion

Overall, the suggested DeepSurv-FBP-ViT framework has shown great promise as a clinically relevant strategy for predicting renal disease and modelling renal survival. The model is able to accurately predict risk and provide high diagnostic accuracy with regularized Filtered Back Projection (FBP) for noise-suppressed image reconstruction, Vision Transformer (ViT) for robust feature extraction and DeepSurv for survival prediction. Experimental results indicate that the proposed framework achieves the best accuracy and interpretability compared to the CNN-based and Cox proportional hazards models with respect to different evaluation metrics. By combining imaging variables with clinically meaningful variables, the approach allows for the identification of patients early, customization of treatment planning, and improved patient outcomes. Higher multimodal data fusion, real-time clinical deployment and larger multi-institutional datasets will improve generalizability and usability of the framework in the future for application in the practical context.

Declaration

Data Availability: Data will be made available on request.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

Ethical Approval: The declaration is "Not Applicable".

Reference

[1] Yoshihisa Miyamoto et al., "Association of Chronic Kidney Disease with Total and Site-Specific Cancer Incidence in Participants of the Japan Public Health Center-based Prospective Study," *Nephrology Dialysis Transplantation*, vol. 38, no. 6, pp. 1487-1496, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[2] Paul Cockwella, and Lori-Ann Fisher, "The Global Burden of Chronic Kidney Disease," *The Lancet*, vol. 395, no. 10225, pp. 662-664, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[3] N. Sasikaladevi, and A. Revathi, "Digital Twin of Renal System with CT-Radiography for the Early Diagnosis of Chronic Kidney Diseases," *Biomedical Signal Processing and Control*, vol. 88, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[4] Md Nazmul Islam et al., "Vision Transformer and Explainable Transfer Learning Models for Auto Detection of kidney Cyst, Stone and Tumor from CT-Radiography," *Scientific Reports*, vol. 12, no. 1, pp. 1-14, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[5] Kiran Kumar Patro et al., "Application of Kronecker Convolutions in Deep Learning Technique for Automated Detection of Kidney Stones with Coronal CT Images," *Information Sciences*, vol. 640, pp. 1-16, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[6] Nidhi Dwivedi et al., "Myofibroblast Depletion Reduces Kidney Cyst Growth and Fibrosis in Autosomal Dominant Polycystic Kidney Disease," *Kidney International*, vol. 103, no. 1, pp. 144-155, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[7] Djeane Debora Onthoni et al., "Latent Space Representation of Electronic Health Records for Clustering Dialysis-Associated Kidney Failure Subtypes," *Computers in Biology and Medicine*, vol. 183, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[8] T.D.K.S.C. Gunasekara et al., "Urinary Biomarkers Indicate Pediatric Renal Injury among Rural Farming Communities in Sri Lanka," *Scientific Reports*, vol. 12, no. 1, pp. 1-13, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[9] Andreas S. Panayides et al., "AI in Medical Imaging Informatics: Current Challenges and Future Directions," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 7, pp. 1837-1857, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[10] Anas Abu-Doleh, and Amjed Al Fahoum, "XgCPred: Cell Type Classification using XGBoost-CNN Integration and Exploiting Gene Expression Imaging in Single-Cell RNAseq Data," *Computers in Biology and Medicine*, vol. 181, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[11] G. Loganathan, and M. Palanivelan, "An Explainable Adaptive Channel Weighting-based Deep Convolutional Neural Network for Classifying Renal Disorders in Computed Tomography Images," *Computers in Biology and Medicine*, vol. 192, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[12] Medi Hunter, Kidney Colorized CT (Normal, Cyst, Tumor, Stone), Kaggle, 2025. [Online]. Available: https://www.kaggle.com/datasets/shuvokumarbasakbd/kidney-colorized-ct-normal-cyst-tumor-stone?utm_source=chatgpt.com

[13] Karthikadevi M, and Sneha N, "Early Chronic Kidney Disease Detection using Machine Learning Techniques," *2025 International Conference on Machine Learning and Autonomous Systems (ICMLAS)*, Prawet, Thailand, pp. 151-155, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[14] Dina Saif, Amany M. Sarhan, and Nada M. Elshennawy, "Early Prediction of Chronic Kidney Disease based on Ensemble of Deep Learning Models and Optimizers," *Journal of Electrical Systems and Information Technology*, vol. 11, no. 1, pp. 1-31, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[15] Sulima M. Awad Yousif et al., "Early Detection of Chronic Kidney Disease using Eurygasters Optimization Algorithm with Ensemble Deep Learning Approach," *Alexandria Engineering Journal*, vol. 100, pp. 220-231, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[16] Md. Ariful Islam, Md. Ziaul Hasan Majumder, and Md. Alomgeer Hussein, "Chronic kidney Disease Prediction based on Machine Learning Algorithms," *Journal of Pathology Informatics*, vol. 14, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[17] S. Yogalakshmi, and M. Anitha, "A Hybrid Deep Learning Approach for Chronic Kidney Disease Detection using Convolutional Neural Networks and Support Vector Machines," *2025 International Conference on Inventive Computation Technologies (ICICT)*, Kirtipur, Nepal, pp. 620-625, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[18] K. Ramu et al., "Hybrid CNN-SVM Model for Enhanced Early Detection of Chronic Kidney Disease," *Biomedical Signal Processing and Control*, vol. 100, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[19] Xi Tan et al., "Deep Filtered Back Projection for CT Reconstruction," *IEEE Access*, vol. 12, pp. 20962-20972, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[20] Lin Hao et al., "Deep Learning-based Survival Analysis for High-Dimensional Survival Data," *Mathematics*, vol. 9, no. 11, pp. 1-18, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[21] MD Nazmul Islam, CT Kidney Dataset: Normal-Cyst-Tumor and Stone, Kaggle, 2021. [Online]. Available: https://www.kaggle.com/datasets/nazmul0087/ct-kidney-dataset-normal-cyst-tumor-and-stone

[22] Orvile, T2-weighted Kidney MRI Segmentation, Kaggle. [Online]. Available: https://www.kaggle.com/datasets/orvile/t2-weighted-kidney-mri-segmentation

[23] Siat Syx, CT2USforKidneySeg, Kaggle, 2021. [Online]. Available: https://www.kaggle.com/datasets/siatsyx/ct2usforkidneyseg

[24] Hisham Abdeltawab et al., "A Pyramidal Deep Learning Pipeline for Kidney Whole-Slide Histology Images Classification," *Scientific reports*, vol. 11, no. 1, pp. 1-9, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[25] Muneer Majid et al., "Enhanced Transfer Learning Strategies for Effective Kidney Tumor Classification with CT Imaging," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 8, pp. 421-432, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[26] Srisopitsawat Pavarut et al., "Improving Kidney Tumor Classification with Multi-Modal Medical Images Recovered Partially by Conditional CycleGAN," *IEEE Access*, vol. 11, pp. 146250-146261, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[27] Erdal Özbay, Feyza Altunbey Özbay, and Farhad Soleimanian Gharehchopogh, "Kidney Tumor Classification on CT Images using Self-Supervised Learning," *Computers in Biology and Medicine*, vol. 176, 2024. [CrossRef] [Google Scholar] [Publisher Link]