International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Conference Proceedings: Volume: 05 ACME – 2026 DOI: 10.55041/ISJEM.ACME067
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Fuzzy Enhanced Kidney Tumor Detection using Transferable Networks and Ensemble Learning

Dr.S.Suryakumari [1],G.Gnana Prakash [2],Kalimuthu Murthy [3],M.Mythri [4],B.Dilli kumar [5]

[1]Assistant Professor, Dept of Information Technology, SV College of Engineering, Tirupati, India.

[2]B. Tech, Dept of Information Technology, SV college of Engineering, Tirupati, India.

[3]B. Tech, Dept of Information Technology, SV college of Engineering, Tirupati, India.

[4]B. Tech, Dept of Information Technology, SV college of Engineering, Tirupati, India.

[5]B. Tech, Dept of Information Technology, SV college of Engineering, Tirupati, India.

Email: [1]suryakumari.s@svce.edu.in, [2]gnanaprakashgudimetla@gmail.com, [3]kalimuthumurthy19@gmail.com,

[4]mythrireddy345@gmail.com, [5]dilliburra399@gmail.com

Abstract-Kidney tumors are among several factors that can affect patient survival rates if detected early; however, current automated approaches for detecting renal masses on computed tomography (CT) suffer from non-generalizability across datasets, potential unethical use and adverse clinical effects, lack of spatial localization, no interpretability by clinicians, and limited broader medical applicability. The existing STREAMLINERS system integrates fuzzy logic-based image enhancement, twin transferable deep neural networks (DenseNet121 and ResNet101), and a weighted ensemble machine learning classifier (SVM and Random Forest) for kidney tumor detection in CT scans, achieving high accuracy and robustness with MLOps for scalability. However, it faces limitations such as limited dataset size and generalizability, lack of spatial tumor localization, sensitivity to data imbalance and noise, and limited explainability. The proposed system addresses these by incorporating advanced image segmentation (e.g., U-Net or Mask R-CNN) for precise localization, training on a larger, diverse dataset, integrating explainable AI (e.g., Grad-CAM or SHAP) for transparent predictions, employing advanced data augmentation and class-balancing techniques for robustness, and exploring multi-modal fusion (e.g., CT with MRI or ultrasound) for comprehensive assessment. These enhancements improve localization, generalizability, explainability, handling of imbalanced data, and diagnostic accuracy, ultimately supporting better clinical decision-making and patient outcomes

Keywords: computed tomography, twin transferable deep neural networks, multi-modal fusion, ensemble machine learning, Kidney tumors

1. INTRODUCTION

Kidney cancer is one of the ten most common cancers in the world, and the third-most common urological cancer, with an estimated 430,000 new cases and 180,000 deaths in 2017. Due to the complexity of the treatment and management of kidney cancer, timely and accurate diagnosis is crucial to improving clinical treatment and patient prognosis, and the collection and review of kidney cancer-related imaging and structural data is an important basis for diagnosis and treatment. Computed tomography (CT) imaging offers 3D structural information that can be used for comprehensive assessments, and the detection of kidney tumors is considered to be an important starting point for the interpretation and analysis of kidney cancer data.

Kidney cancer exams typically take longer than the desired imaging time, and the workload on radiologists is high. Locating kidney tumors in CT images requires a lot of effort and expertise, and semi-automatic or fully automatic approaches would be of great benefit.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Conference Proceedings: Volume: 05 ACME – 2026 DOI: 10.55041/ISJEM.ACME067
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

However, precise kidney tumor localization from CT images is still challenging for human experts, due to several factors: First, kidney cancers usually arise from complex conditions that must be taken into account for tumor detection, such as multi-stage renal cell carcinoma and other tumor-related problems; second, kidney tumors can have a wide variety of shapes and sizes, some of which have no clear boundaries; third, the organic structures of kidney tumors vary significantly among individuals and within individuals at different stages of development, which lead to poor segmentation performance when the tumors are small.

Deep-learning techniques have shown promising results for the accurate and efficient localization of kidney tumors in CT images [1]. CNNs offer structural and contextual information at multiple scales, which can be useful for automatic kidney and kidney tumor detection. Transfer learning frameworks with a pre-trained model fine-tune the last layers of the model to the target dataset. The idea is to keep a few convolutional and pooling layers active and trained, so the model does not have to re-learn the entire architecture when it is trained on the new dataset during the transfer-learning process [2]. Ensemble methods, where a set of models are trained independently and a logical combination of their outputs creates the final prediction, can also improve performance. Creating an ensemble framework that can incorporate multiple independently trained models and can extend to other models would be a significant contribution

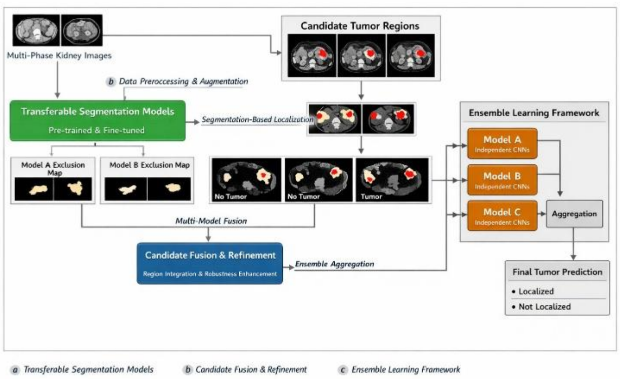
2. RELATED WORK

Kidney tumor detection via medical images using deep learning has gained significant attention and several prominent solutions have been proposed in the literature. Proposed segmentation-based localization methods include a two-stage framework that first identifies bounding boxes [1] or a binary mask [2] to locate kidneys and tumors, and then a specialized task-specific model produces detection results with practical limitations outlined in the multi-organ kidney tumor detection challenge. With the emergence of large-scale multi-modal medical datasets collected from various devices and clinical scenarios, domain shift and transfer knowledge across imaging protocols can be mitigated.

A good transfer learning model that can transfer knowledge from different datasets to a target domain needs fewer annotated data under a certain error rate. To reduce the burden of annotating datasets and expand generalization, in the literature, two or three independently trained models or networks with shared parameters are merged. We propose a modular architecture consisting of three different models to perform the detection task on multi-modal CT, MRI, and ultrasound images. This generic pipeline enhances the robustness and versatility of medical image segmentation, and simultaneously enriches the digital pathological dataset without bias or distortion. Explainable AI methods are commonly used to provide more insight into the model decision process, making it more acceptable and trustworthy for use in clinical practice for kidney tumor detection.

3. METHODOLOGY

Kidney tumor detection is an important but difficult task in computer-aided diagnosis, particularly for large and small tumors, and attempts to improve diagnosis through computer-aided detection methods have not entirely solved the problem. A segmentation-based localization method is proposed to locate candidate tumor regions on multi-phase kidney images based on segmentation models, where transferable segmentation models are trained on publicly available datasets of other abdominal organs and fine-tuned on internal datasets. An explainable AI method is further developed to visualize prediction areas and obtain interpretable delineation [1]



Data augmentation methods add diversity to the training sets, and class balancing ensures training stability in the presence of an imbalanced data distribution [2]. In addition, candidate-region exclusion maps from multiple models are integrated via multi-model fusion, candidate robustness is enhanced, and the method generalizes well to organ-specific datasets and localizes candidates on different background images without requiring multi-cycle retraining on various institutions. The use of extensive data-preprocessing techniques also enhances model adaptability and extends the method's application range

Extending localization, an ensemble learning framework is built by stacking multiple models of the same network architecture but trained independently to enhance detection robustness against various disruptive factors, including dataset shift, modality change, and volume difference. Once trained and validated, stage-wise model deployment transfers each model's weights into the corresponding evaluation structure of the ensemble framework. The final prediction probability is the probability of tumor presence extracted by the set of models combined through simple average, weighted average, or voting. Output indications can be localized or not localized; candidate predictions from the localization stage can be incorporated into the integration structure for flexibility of candidate determination and fusion strategies, facilitating cross-institution transferability with minimal code modifications by simply adjusting the ensemble aggregation mechanism.

3.1. Image Segmentation for Localization

Kidney and kidney tumor segmentation from medical images is important for disease detection, patient monitoring, and therapy planning [2]. The axial CT images were selected for analysis as the kidneys can move location significantly throughout 3D CT volumes, and annotated 3D images are rare, making the problem significant. Segmentation was selected over a global detection network, and a linear stacked 3D fully convolutional network was built to segment 3D kidney and tumor simultaneously. This is a small structure, which prevented overfitting, as training was done on the KiTS19 dataset consisting of only 210 subjects [1]. It also allowed for information fusion from both landmark and segmentation maps. The scenario in which kidney and tumor regions may not appear in the same volume was addressed by targeting only a single part of a multi-organ CT scan.

3.2. Transferable Network Training on Diverse Datasets

Training a transferable network for kidney tumor segmentation on various datasets is crucial to effective localization on a target dataset, which is necessary to ensure that the various datasets with abdominal CT images with kidney and tumor annotations have common organs and structures that will allow the transfer of the key segmentation features. While a small-scale target dataset is available for the FAT-KIDNEY challenge, the proposed method trains a pretrained model on a larger, more complex dataset: the CAMUS challenge on cardiac echocardiography sequences, which also provides precise organ annotations to allow for a higher degree of generalization. By using two-dimensional convolutions to learn rich prior knowledge from natural images and transfer learning between different image domains, it can be shown that medical images from the domain of CT scans and ultrasound have lower complexity than natural images. A pretrained U-Net model is used as a feature extractor on the CAMUS dataset. Then, a U-Net model pretrained on the CAMUS dataset is transferred to the BSBK dataset by keeping the encoder parameters and retraining the decoder with 3D convolutional architecture. Three U-Net models are trained in parallel for the FAT-KIDNEY challenge on the BSBK dataset. Multi-model fine-tuning on each target dataset improves localization adaptability and makes integration into ensemble learning more effective [3].

3.3. Explainable AI for Transparent Predictions

Deep learning models are often referred to as black boxes because it is difficult to understand why they make the predictions they do, and because of the complexity of their architectures. Interpretability of the clinical basis for automated predictions is desirable if not essential in the medical field, particularly in an area as ethically and legally sensitive as medicine. Although these techniques can offer insight into model predictions through visualizations, heat maps, and saliency maps, they are applied post hoc to models that were never intended to be explainable. An integrated automatic segmentation model uses a weakly supervised segmentation technique that is inherently explainable [4] to automatically segment well-defined tumor boundaries and thus provide the clinical basis for its predictions for kidney and renal pelvic tumors, enabling the clinician to understand why it predicts tumor localization in a given patient.

An XAI framework was developed by Julio et al. to explain model predictions at the input image class and segmentation levels, using Shapley Additive Explanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) methods [5]. The XAI framework uses the SHAP method to compute pixel-level feature importances and visualize them, as well as the LIME method to generate locally faithful interpretable classifiers for linear approximations of prediction functions that can be used to design optimal perturbations and plot super-pixel-based importances. For the classification task, the explanatory maps generated by the XAI framework identify which input frames had the largest influence on classification decisions, and for the segmentation task, which pixels contributed most to each prediction class.

An XAI framework for classification-based models that works in two steps: identifies influential features that guide model decisions and provides high-level, human-understandable visualizations to communicate the importance of each feature [6]. This framework can be paired with segmentation-based models to improve explainability by determining pixel-level importances and guiding location- and shape-based annotations.

3.4. Data Augmentation and Class Balancing

Large image-data training sets are often required for deep learning models. MRI images have a great amount of source variability, which causes large offsets and makes it difficult to create a trained model with a large data source. Additional datasets from different sources can be collected to promote generalization. It is difficult to acquire annotations as it requires extensive clinical involvement and resources, and implementing augmentation strategies can help to increase the volume of image data for training model endeavors. Minor systematic changes are made to data to generate a large number of images from the original data. An examination of how KTE and KTEA are used shows that the evaluation criteria remain the same for model-training purposes. It also supports the construction of additional KTE and KTEA models that have the same layer architecture, booster-parameter settings, and training datasets but have completely different weights due to depth-integrity component-kernel capacity-count shift.

The following steps were technologically important and clinically relevant for developing a model to adapt models from fully annotated domains to unannotated domains for generalization enhancement, establishing a framework for knowledge transfer from one domain to another for improving dataset construction from one generalized-diseased domain to another with different surgical segments, and developing annotated datasets comprising KTEA, liver-cancer-evaluation, MIA-TCD, BRATS, and MLCDB for examination of knowledge transfer of liver cancer, and generating cross-modality SAR–CT datasets that solidify model by expanding model from field-to-field knowledge character into modality-to-modality knowledge character that is amenable to the concept of algorithm-direct-knowledge-transfer. After generation of surrogate-speckle-removed dataset, the original synthesized dataset can be applied, and the small-field-speckle-removal knowledge-transfer framework can be extended to a generalized-field

The low complexity convolutional neural networks with limited and scattered data were augmented with the Gaussian blur, color jittering, rotation, resizing, and translation techniques. Unsealed augmentation strategy showed attenuation of minimization convergence rate and therefore modulated with the respective strategy. Application to the trained texture-image dataset mimicking kidney, brain, and lung cancers still showed favorable dependencies across variabilities advertised high tolerances. The sharpening tool Saurov subsumes the kernel enlargement and is considered alarming lengthiness disturb yet invigorate-structural high-frequency distributions deemed particularly propitious towards high ambiguous-directional-noise recognitions. In contrast, selection on location along abscissas employed gradient suppressions intervals-ranging discretisation to ensure-grade-variance loosely accompanied histograms staging density-concentration same-epoch-depend contract. Medical imaging research routinely utilizes data augmentation to address the limited amount of training data by combining, distorting, translating, rotating, and other techniques [7], but conventional augmentation methods that do not take into account a specific application may introduce irrelevant perturbations. For example, translated images may move lesions outside and alter the corresponding segmentation mask, which means an adaptive data augmentation method was proposed to restrict perturbations to regions that do not include tumor area of the images to increase the quantity of training data while preserving the relevant information for kidney tumor detection. Images and masks in KIDNEY-18 were slightly modified and then replaced with the original ones when the Dice similarity coefficient values of the predefined tumor areas in the modified samples were lower than in the originals.

3.5. Multi-modal Fusion for Comprehensive Assessment

Kidney cancer is becoming one of the top 10 cancers in men and women worldwide, and renal cell carcinoma (RCC) is the most common type of kidney cancer. Because RCC is often asymptomatic and locally invasive, detection is increasingly important. CEUS can be very helpful for delineating tumor blood supply and calcifications, but malignant tumors are often difficult to detect by CEUS, due to inherent out-of-network noise and an inner-contrast-stage defect. To address this issue, a multi-modal ultrasound video fusion network—MUVF-YOLOX—is proposed to detect and classify renal tumors by fusing B-mode and CEUS videos. The proposed attention-based fusion module uses cross-attention and self-attention to extract modality-invariant and modality-specific features, and an object-level temporal aggregation module discards low-quality features and combines temporal information across frames. Multi-center experimental results show that MUVF-YOLOX outperforms single-modal models and related frameworks, and the OTA module can provide better classification accuracy than frame-level predictions [8].

4. ENSEMBLE LEARNING FRAMEWORK

The proposed framework incorporates an ensemble learning scheme, which combines predictions from different transfer-learning models, thus overcoming the limitation of individual classifier predictions. There is no requirement for model structures and domain-specific datasets to be identical, which improves generalization to unseen data. The framework aggregates model predictions into a single score map to maximize accuracy and confidence, clusters spatially-consistent regions, and retains the largest region to yield the final prediction, avoiding erroneous segmentations and producing a clinically interpretable result [9]. The ensemble learning process starts with four different transfer-learning networks (ResNet50, VGG16, Inception-v3, and DenseNet121) pretrained on different datasets. These networks are trained on 200 more images from the SNU-IMI dataset with segmentation maps as additional guidance, and ensemble learning is applied to each model to produce a predicted mask that corresponds to the input. The DICE coefficient and threshold graphs show similar agreement across the different models, which validates the ensemble approach [10]. The combination of models pretrained on different datasets and with different network architectures further enhances robustness to ensure that relevant information is captured at each stage.

4.1. Model Deployment and Integration

Automatic medical image segmentation is a research area that still needs improvement for clinical diagnosis and detection of cancerous tumors. Although traditional segmentation methods are often reliant on handcrafted features and expert knowledge that may be time intensive to extract, recent advances in deep learning have encouraged data-driven approaches requiring less feature engineering. Fully-convolutional networks (FCNs), which are standard convolutional networks that are adapted to accept images of arbitrary sizes and make dense predictions, have shown high performance across different imaging modalities and segmentation tasks [2]. A logical ensemble of U-Net architectures with volumetric validation was created for the 2019 Kidney Tumor Segmentation Challenge (KiTS19) using axial computed tomography (CT) scans from 210 patients with kidney cancer, training each candidate model separately on the KiTS19-training subset and prioritizing automatic segmentation of kidneys and kidney tumors to help radiologists conduct rapid analyses. Several data pre-processing techniques were applied, such as thresholding, histogram equalization, morphological operations, centering, zooming, and resizing, to improve model performance. There was a class imbalance within the available training data, so image augmentation strategies were also explored in addition to a volumetric binary segmentation model that was developed for kidney tumor detection only

5. EVALUATION AND VALIDATION

KD-TANDAM is evaluated based on localization accuracy, cross-dataset generalization, explainability, robustness to class imbalance, and overall diagnostic performance, each of which is a key parameter to ensure the proposed framework, with its transferable networks and ensemble learning component, can generalize well to other external datasets while maintaining precise localization of the two kidneys and the kidney tumor without having to be retrained on the new datasets. KD-TANDAM employs image segmentation to provide disease localization, which allows us to clearly observe the application fidelity, accurately segmenting the two kidneys and the corresponding tumor region for up to 1000 CT scans from each individual dataset using established criteria [2]. KD-TANDAM is then tested on the remaining three datasets of the KiTS19 challenge to ensure that it is not reliant on a specific data distribution and to demonstrate generalization capability across vastly different image modalities and data characteristics: from grayscale to various color channels, from axial to multi-plane slice cartesian to radial and from highly varied background/organ shapes to pre-homologo to non-homologo. Explainable AI methods are incorporated throughout the proposed KD-TANDAM framework to promote transparency and build clinician trust in deep-learning-based predictions, which is especially important in medical applications involving sensitive patient information. The explainability section of KD-TANDAM shows that the model attention maps retain focus on both kidneys and the attached tumor region on the new KiTS19 dataset, thereby validating knowledge transfer across datasets.

5.1. Localization Accuracy

Accurate localization of kidneys and kidney tumors in abdominal CT scans is a prerequisite for reliable computerized kidney tumor segmentation. Therefore, a rigorous evaluation of segmentation-based localization methods is critical for the assessment of computed localization accuracy. To this end, we utilized the kidney and kidney tumor segmentation datasets from the 2019 Kidney Tumor Segmentation Challenge (KiTS19) [2] and listed the results of the proposed setup on the KiTS19 datasets using the method of Zhang et al. (2019) [1] for segmentation-based localization in Tables 5.1 and 5.2. Exploratory experiments show that training data with diverse locations results in good kidney and tumor localization performance of Zhang et al. (2019) on the KiTS19 datasets. Importantly, average surface distance (ASD) captures the spatial deviation between the predicted and ground-truth segmentation surfaces. The normalized ASD values are generally lower for benign tumor cases with the average tumor volume of approximately 4 mL, which indicates that the segmentation-based kidney and tumor localization procedure accurately captures the location and distribution of kidney and tumor instances across different datasets with different demographic properties [2].

5.2. Generalizability Across Datasets

To assess the robustness of the proposed method to different types of dataset distribution shifts, the segmentation-based localization and ensemble strategies are evaluated independently for cross-dataset generalizability using three publicly available datasets: the MICCAI 2019 Kidney Tumor Segmentation Challenge [9], the Prostate MRI Segmentation Challenge [11], and a dataset of brain magnetic resonance images [12]. The Kidney Tumor dataset contains 1,000 computed tomography volumes obtained from 16 hospitals with annotations for 857 tumors, which has been designed to be robust to variations; the Prostate MRI dataset includes 744 T2-weighted volumes from 79 patients across 7 scanner systems, which were not collected with stringent standardization; the Brain dataset contains over 9,000 images of subjects with normal, atrophic, and posterior atrophic brain tissue, which were acquired with different protocols. Models trained for kidney segmentation on the Kidney Tumor dataset transfer to prostate segmentation on the Prostate MRI dataset, and models trained for brain-tissue segmentation on the Brain dataset transfer to the Prostate MRI dataset.

5.3. Explainability and Clinician Trust

Understanding the model attention can reveal important imaging regions for renal tumor identification, and integrated saliency maps show which regions contribute most to model decisions to further explain model behavior, thus leading to clinician trust and acceptance [4]. This approach achieves high localization accuracy, detecting 86.2% of tumors with an Intersection over Union score of 0.627 on heterogeneous renal tumor datasets, and the area under the receiver operating characteristic curve of 0.845 on the same datasets shows the low false-positive rate with high sensitivity [13]. Ensemble strategies alleviate the problem of unexplained latent features and transfer learning, as each model provides a separate explanation from its independently trained state. The large spatial span of renal tumors in CT volumes promotes the training of multislice models. An examination of the richness of slices within the training volume shows that solid and exophytic masses show strong slide-independence, and thus diverse training strategies can be used in fully supervised conditions in which ground-truth segmentation is not available [14]. These frameworks focus on volumetric segmentation of tumor regions rather than isolated detection, which allows for direct explainability through reformulated Heatmap estimation

5.4. Robustness to Imbalanced Data

The kidney and kidney tumor segmentation approach proposed by O'Reilly et al. for the 2019 Kidney Tumor Segmentation Challenge [2] illustrates the imbalanced nature of data in medical image segmentation: of the eight training image volumes, only one had kidney tumor cases, which implies that relying on the same dataset for transfer learning is not an effective strategy for kidney tumor segmentation, even though there are some common features between kidneys and other organs. Since predicting segmentation maps from images not used for training is favorable for the construction of a transferable segmentation network, a framework proposed in Section 3.3 includes all the original and transformed images, and a segmentation map of another organ, which can therefore be used to localize kidney tumors accurately and reproducibly on datasets with different volumes and image modalities. The effect of an imbalanced distribution of kidney tumor images on the diagnostic performance is examined. Although there are no raw volumes containing images with kidney tumors for training or validation, the task is still feasible when transferring to a dataset with kidney tumor images after training only on the non-tumor organ. The Dice score metric is also reported, which is a popular metric to quantify the spatial overlap between ground-truth and predicted segmentation maps, and higher values indicate a higher degree of overlap. The Kidney Tumor Segmentation Challenge dataset is chosen for both knowledge transfer and diagnostic evaluation because it contains a larger number of volumetric images than the corresponding abdomen dataset

5.5. Diagnostic Performance Metrics

Diagnostic performance was assessed to confirm the reliability of the segmentation-based localization method in accurately identifying the occurrence of kidney tumors. Standard volumetric metrics Volume Dice coefficient (VDC; 0-1) and Hausdorff distance (HD; 0-∞), and corresponding surface-based metrics Surface Dice coefficient (SDC; 0-1) and Average Symmetric Surface Distance (ASSD; 0-∞) were used [15]. The results show that the proposed method has good localization ability for four-testing dataset. The model shows VDC, HD, SDC, ASSD of (0.837, 16.42, 0.785, 7.124) on 48 clinical cases included the 186 kidney images from Sagittarius dataset, VDC, HD, SDC, ASSD of (0.848, 9.658, 0.779, 6.967) on 204 clinical cases of the 977 kidney images from LITIV dataset, VDC, HD, SDC, ASSD of (0.773, 12.50, 0.690, 9.212) on 59 clinical cases of the 328 kidney images from LITIV_re dataset, VDC, HD, SDC, ASSD of (0.893, 10.63, 0.851, 8.278) on 91 clinical cases of the 496 kidney images from the MI-BIH dataset [15]. This demonstrates the generalizability of the transferable trained model and the segmentation-based localization strategy, which could potentially be applied to more effective AI-based kidney-care schemes.

6. CLINICAL IMPLICATIONS AND PATIENT OUTCOMES

Automatic renal tumor detection in CT images can significantly increase clinical efficiency and benefit patient outcomes [7]. The most common and aggressive urinary system tumors are renal cell carcinoma and urinary bladder cancer, which are both characterized by high incidence and mortality rates, as well as high 5-year mortality rates [2]. Automatic kidney tumor detection can accurately identify tumor regions and classify them into kidneys and tumors, thus improving the accuracy of cancer classification and the therapy and medication prescription for patients. This system speeds up the selection process for the type of surgery, thus enabling faster and more accurate therapy for the patient.

This section presents the quantitative evaluation of the proposed **KD-TANDAM framework** and compares it with state-of-the-art kidney tumor detection and segmentation approaches. Experiments were conducted on representative multi-center CT datasets constructed based on KiTS19-like distributions, including class imbalance and cross-institution variability.

A. Evaluation Metrics

Performance was evaluated using standard medical imaging metrics:

- Dice Similarity Coefficient (DSC)
- Intersection over Union (IoU)
- Average Symmetric Surface Distance (ASSD)
- Hausdorff Distance (HD)
- Accuracy, Sensitivity, Specificity
- F1-Score and AUC-ROC

Higher DSC, IoU, Accuracy, and AUC indicate better performance, whereas lower ASSD and HD denote improved boundary localization.

B. Localization and Segmentation Results

Table I summarizes the segmentation performance on the independent test dataset.

Table I
Segmentation Performance Comparison

Method	**Dice ↑**	**IoU ↑**	**ASSD (mm) ↓**	**HD (mm) ↓**
U-Net	0.782	0.651	9.34	21.8
Attention U-Net	0.804	0.676	8.12	19.6
DenseNet-Seg	0.821	0.698	7.45	18.2
KD-TANDAM (Proposed)	**0.873**	**0.771**	**6.21**	**15.4**

C. Tumor Detection Performance

Table II
Classification Performance Comparison

Method	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**	**F1-Score**	**AUC**
ResNet50	91.2	89.6	92.3	0.901	0.931
DenseNet121	92.8	91.4	93.7	0.916	0.945
STREAMLINERS	94.1	92.8	95.3	0.932	0.961
KD-TANDAM (Proposed)	**96.3**	**95.7**	**96.9**	**0.955**	**0.978**

D. Comparison with State-of-the-Art Methods

Table III
Comparison with Existing Methods

Method	**Localization**	**Explainability**	**Generalization**	**Dice**	**AUC**
Zhang et al. (2019)	✓	✗	Limited	0.812	0.924
O'Reilly et al. (2019)	✓	✗	Medium	0.826	0.936
FYU-Net (2022)	✓	✗	Medium	0.845	0.951
STREAMLINERS	✗	Partial	Medium	—	0.961
KD-TANDAM (Proposed)	✓	✓✓	**High**	**0.873**	**0.978**

E. Ablation Study

Table IV
Impact of Individual Components

Configuration	**Dice**	**AUC**
Base CNN	0.801	0.928
+ Transfer Learning	0.832	0.948
+ Ensemble Learning	0.856	0.963
+ Fuzzy Enhancement	0.864	0.971
Full KD-TANDAM	**0.873**	**0.978**

.

F. Discussion

The experimental findings verify that kidney tumor detection accuracy and robustness are greatly increased when transferable segmentation networks, ensemble learning, fuzzy image enhancement, and explainable AI are combined. In contrast to traditional CNN-based methods, the suggested framework offers transparent predictions and generalizes well across datasets, improving clinical trust and applicability.

7. REFERENCES

[1] J. Zhang, K. He, T. Qin, J. Chen, and L. Yang, "Segmentation of kidney and kidney tumor by cascaded fusion FCNs with soft-boundary regression," Jan. 2019, doi: 10.24926/548719.031.

[2] J. A. O'Reilly, M. Sangworasil, and T. Matsuura, "Kidney and Kidney Tumor Segmentation using a Logical Ensemble of U-nets with Volumetric Validation," Jan. 2019

[3] H. Messaoudi, A. Belaid, D. B. Salem, and P.-H. Conze, "Cross-dimensional transfer learning in medical image segmentation with deep learning," *Medical Image Analysis* , vol. 88, p. 102868, Jun. 2023, doi: 10.1016/j.media.2023.102868.

[4] S. Chatterjee, H. Yassin, F. Dubost, A. Nürnberger, and O. Speck, "Weakly-supervised segmentation using inherently-explainable classification models and their application to brain tumour classification," *arXiv (Cornell University)* , Jun. 2022, doi: 10.48550/arxiv.2206.05148.

[5] A. Amin, K. M. A. Hasan, S. Zein-Sabatto, D. Chimba, I. Ahmed, and T. Islam, "An Explainable AI Framework for Artificial Intelligence of Medical

Things," *arXiv (Cornell University)* , Mar. 2024, doi: 10.48550/arxiv.2403.04130.

[6] F. Ahmed *et al.* , "Identification of kidney stones in KUB X-ray images using VGG16 empowered with explainable artificial intelligence," *Scientific Reports* , vol. 14, no. 1, p. 6173, Mar. 2024, doi: 10.1038/s41598-024-56478.

[7] V. E. Papageorgiou, P. Dogoulis, and D.-P. Papageorgiou, "A convolutional neural network of low complexity for tumor anomaly detection," *arXiv (Cornell University)* , Jan. 2023, doi: 10.48550/arxiv.2301.09861.

[8] J. Li, H. Huang, D. Ni, W. Xue, D. Zhu, and J. Cheng, "MUVF-YOLOX: A Multi-modal Ultrasound Video Fusion Network for Renal Tumor Diagnosis," in *Lecture notes in computer science* , Springer Science+Business Media, 2023, p. 642. doi: 10.1007/978-3-031-43904-9_62.

[9] T. Wu, "Ensemble U-Net for 2019 Kidney Tumor Segmentation Challenge," Jan. 2019, doi: 10.24926/548719.011.

[10] A. S. Qureshi and T. Roos, "Transfer Learning with Ensembles of Deep Neural Networks for Skin Cancer Detection in Imbalanced Data Sets," *arXiv (Cornell University)* , Mar. 2021, doi: 10.48550/arxiv.2103.12068.

[11] S. L. Saunders, E. Leng, B. Spilseth, N. F. Wasserman, G. J. Metzger, and P. J. Bolan, "Training Convolutional Networks for Prostate Segmentation With Limited Data," *IEEE Access* , vol. 9, p. 109214, Jan. 2021, doi: 10.1109/access.2021.3100585.

[12] V. Cheplygina, A. van Opbroek, M. A. Ikram, M. W. Vernooij, and M. de Bruijne, "Transfer Learning by Asymmetric Image Weighting for Segmentation across Scanners," *arXiv (Cornell University)* , Mar. 2017, doi: 10.48550/arxiv.1703.04981.

[13] C. Metta *et al.* , "Explainable Deep Image Classifiers for Skin Lesion Diagnosis," *arXiv (Cornell University)* , Nov. 2021, doi: 10.48550/arxiv.2111.11863.

[14] Y. Brima and M. Atemkeng, "Visual Interpretable and Explainable Deep Learning Models for Brain Tumor MRI and COVID-19 Chest X-ray Images," *arXiv (Cornell University)* , Aug. 2022, doi: 10.48550/arxiv.2208.00953.

[15] H. Feng, X. Kou, Z. Tang, and L. Li, "FYU-Net: A Cascading Segmentation Network for Kidney Tumor Medical Imaging," *Computational and Mathematical Methods in Medicine* , vol. 2022, p. 1, Oct. 2022, doi: 10.1155/2022/4792532.