A Federated Learning Model for Privacy-Preserving and Cross-Domain Kidney Stone Detection in Medical Imaging

Lucas Sotomaior, Luiz F. B. Fonseca, Matheus A. C. Zangari
Rodrigo K. Krebs, Alceu Britto Jr, and Eduardo K. Viegas

***Abstract*— Kidney stones significantly impact healthcare systems, with diagnosis typically requiring time-consuming Computed Tomography (CT) scan consultations between physicians and radiologists, often delaying patient care. Achieving a quick and accurate diagnosis is essential to ensure timely and effective treatment, which has motivated the development of Deep Neural Network (DNN)-based approaches for automated kidney stone detection. However, building effective models remains challenging, as it often requires access to large and diverse datasets that are siloed across institutions, and sharing such medical data is rarely feasible due to strict privacy regulations and patient confidentiality concerns. This paper proposes a privacy-preserving Federated Learning (FL) framework that enables multiple medical institutions to collaboratively train a DNN model without sharing sensitive patient data. Each institution trains a local model on its private dataset, and a centralized trusted server securely aggregates model parameters. We evaluate our approach using abdominal CT scan image datasets from two distinct institutions. Experimental results demonstrate that our proposed model achieves high classification accuracy within the same training environment, with an F1-score of up to 0.94. In addition, in cross-dataset evaluations, our approach outperforms traditional centralized baselines, showing significantly lower performance degradation while preserving patient privacy.**

I. Introduction

The growing digitization of healthcare has resulted in the widespread availability of medical imaging data, driving the development of Machine Learning (ML) models to assist in diagnostics and treatment planning [1]. Among various medical challenges, automatically classifying kidney stones using Computed Tomography (CT) images is critical for guiding appropriate treatment strategies [2], specially in emergency room scenarios. In this direction, an ML model can be trained to detect kidney stones by learning patterns from annotated medical images, such as CT scans, where experts previously label the presence and characteristics of stones. During training, the model adjusts its internal parameters to accurately distinguish between images with and without kidney stones, enabling automated detection considering unseen data. Despite the promise of ML solutions in improving diagnostic accuracy, their implementation often requires sensitive patient data across multiple medical institutions. Such a need raises significant concerns regarding privacy, data protection, and compliance with legal frameworks such as the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA) [3].

[1]Lucas Sotomaior, Luiz F. B. Fonseca, Alceu S. Britto Jr, and Eduardo K. Viegas are with Graduate Program in Computer Science (PPGIa) at Pontifícia Universidade Católica do Paraná (PUCPR), Brazil. Rodrigo K. Krebs and Matheus A. C. Zangari are with Urology Department of the Hospital Nossa Senhora das Graças (HNSG), Brazil. {lucas.sotomaior, luiz.fonseca, krebs.rodrigo, alceu, and eduardo.viegas}@ppgia.pucpr.br

Recent works have demonstrated the effectiveness of Deep Neural Network (DNN) techniques for classifying and diagnosing diseases using medical images [4]. In general, proposed schemes use a centralized learning strategy, aggregating data from multiple sources into a single repository for model training, yielding significantly high detection accuracies [5]. However, despite their success, these approaches rely on the centralization of data, which is often impractical in real-world clinical settings due to privacy, legal, and logistical constraints [6]. Sharing sensitive medical data across institutions can expose patients to privacy breaches, introduce risks of data misuse, and lead to compliance violations with strict regulatory frameworks [7]. Additionally, technical challenges related to data standardization, interoperability, and the secure transfer of large volumes of data further complicate the centralized model training in healthcare environments.

Federated Learning (FL) has emerged as a promising paradigm to address these challenges [8]. In practice, it enables the collaborative training of ML models across multiple decentralized devices or institutions without requiring the exchange of raw data. Instead, only model updates, such as gradients or parameters, are shared, thereby preserving data locality and protecting sensitive information. In the field of medical imaging, FL has been successfully applied to a range of tasks, including brain tumor segmentation [9], histopathological cancer detection [10], and autism spectrum disorder (ASD) identification [11]. These studies often demonstrate that collaborative learning across institutions can produce models with performance comparable to or exceeding that of traditional centralized approaches. Additionally, FL offers further advantages, such as mitigating data silos, improving model generalization by leveraging heterogeneous datasets, and facilitating compliance with stringent data protection regulations.

Unfortunately, despite the progress in applying FL to various medical imaging tasks, its application to kidney stone detection remains largely unexplored. Kidney stone classification presents unique challenges, particularly due to the variability in imaging conditions across different clinical environments [12]. Overcoming domain shifts caused by variations in image acquisition protocols, patient populations, and equipment remains a significant challenge for maintaining consistent model performance. Surprisingly, proposed schemes often overlook domain shifts and assume the use of a single dataset with consistent data collection characteristics [13]. Notwithstanding, when multiple datasets are involved, the privacy challenge of sharing sensitive data across institutions or devices is frequently ignored.

Contribution. In light of this, we propose a novel application of FL to kidney stone classification using CT images. The proposed approach is implemented in two stages. First, each participating institution builds its local model using its private dataset without sharing the raw data externally. This decentralized training process allows institutions to leverage their sensitive medical data to build effective models while maintaining complete control over data privacy and security. Second, local models are aggregated to construct a global model that captures shared knowledge in a cross-domain setting. This design allows us to assess whether FL can achieve performance comparable to centralized learning methods while adhering to strict data protection and privacy regulations. Our central insight is to model kidney stone detection as an FL task, enabling collaborative training across multiple medical institutions while preserving patient privacy.

In summary, the main contributions of this work are as follows:

- We propose a framework for applying FL to the kidney stone classification problem, tailored to the healthcare context with privacy preservation as a central requirement;
- We design and conduct a comprehensive experimental evaluation comparing FL with centralized and individual clinic-specific models, analyzing classification performance and system scalability;
- We identify practical challenges, opportunities, and best practices for deploying FL in real-world medical settings, contributing with insights for researchers and practitioners interested in privacy-aware AI for healthcare.

II. Preliminaries

A. ML for Kidney Stone Detection

Machine Learning (ML), particularly Deep Neural Networks (DNNs), has been successfully applied to detect kidney stones from CT images. Early approaches based on handcrafted features [14] have mainly been replaced by deep learning models capable of learning rich visual representations directly from data [5], [15]. Architectures such as ResNet, VGG, and InceptionV3 have achieved high classification accuracy, often exceeding 95% in controlled environments [5], [16]. However, despite these advances in representation learning, significant challenges remain, particularly in ensuring generalization across data from different CT machines and institutions. The limited availability of annotated images that reflect the full spectrum of variability in imaging protocols, patient populations, and equipment settings restricts model robustness in real-world applications.

B. The challenges of ML for healthcare

As mentioned before, despite significant progress, the application of ML in healthcare faces major challenges. Privacy regulations such as HIPAA and GDPR restrict data sharing, making centralized training strategies impractical in many clinical scenarios [3], [7]. Moreover, variations in imaging protocols, CT equipment, and patient demographics across institutions introduce domain shifts that limit model generalization. These factors hinder the development of robust, scalable models using traditional approaches. Federated Learning (FL) has emerged as a promising alternative, enabling collaborative model training without exposing sensitive data [17], [9]. Shortly, FL may also pave the way for multimodal and multiview solutions, requiring extensive and diverse datasets from multiple sources to model complex diagnostic patterns across institutions while preserving privacy.

III. Related Works

Recent advances in deep learning (DL) have significantly enhanced the automatic detection and classification of kidney stones using computed tomography (CT) images. Centralized approaches leveraging Convolutional Neural Networks (CNNs) and transfer learning techniques have reported impressive results in various studies. For instance, in [15], the authors developed fine-tuned models based on VGG16, ResNet50, Alexnet, and InceptionV3, achieving classification accuracies up to 99.96% through extensive preprocessing and hyperparameter optimization. Similarly, Yildirim et al. [5] proposed a deep learning model using coronal CT images and the XResNet-50 architecture, reporting an accuracy of 96.82% and F1-score of 0.97. Their publicly available dataset has further supported reproducibility and benchmarking in the field. Other efforts, such as those of [16], have proposed hybrid models combining ResNet101 with custom CNNs, achieving near-perfect classification across multiple kidney abnormalities, including stones, cysts, and tumors.

Despite the effectiveness of these centralized methods, their deployment in real-world clinical environments presents notable challenges. Medical imaging datasets are typically siloed across institutions with varying imaging protocols, scanner types, and data quality. Additionally, privacy regulations restrict the centralized aggregation of patient data. These factors hinder the generalizability of models trained on homogeneous datasets and reduce their performance in multi-institutional or real-time clinical settings. For instance, the authors in [18] found substantial drops in model accuracy when testing across different CT planes, with performance in testing scenarios dropping as low as 63% despite high training accuracies above 98%. To address these limitations, recent research has explored federated learning (FL) as a decentralized alternative. FL enables the collaborative training of models across multiple institutions without sharing raw data, thus preserving patient privacy while enhancing model generalizability.

In this direction, the work in [19] describes a federated transfer learning framework tailored to kidney disorder detection. Their approach allows geographically distributed

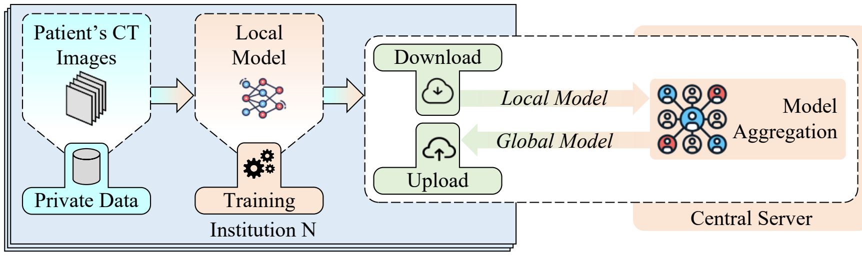


Fig. 1: Proposed privacy-preserving federated learning for kidney stone detection in medical imaging. Each participating institution trains a local model using its private data. A centralized and trusted server then aggregates these locally trained models to construct a global model. .

healthcare centers to train diagnostic models while retaining local data jointly, thus overcoming the constraints of centralized learning architectures [20]. Combining FL with transfer learning ensures that domain-specific knowledge from pre-trained models is adapted efficiently to local data distributions, improving accuracy while respecting data governance regulations. In summary, while centralized deep learning models have demonstrated high classification performance in controlled environments, federated learning provides a promising path forward for deploying AI in heterogeneous and privacy-sensitive healthcare ecosystems.

IV. Privacy-preserving Federated Learning for Kidney Stone Detection in Medical Imaging

To address the aforementioned challenges of preserving privacy in kidney stone detection from medical images, our proposed model formulates the task within an FL framework. Figure 1 shows the implementation strategy of our proposed scheme, which is implemented in two phases.

First, we frame kidney stone detection as a DNN-based image classification task, utilizing CT images to assess the presence of kidney stones. In our approach, a DNN model is trained to learn complex visual features from these images, enabling the automatic identification of kidney stones. This supports clinicians by providing an objective diagnostic tool to enhance accuracy, reduce workload, and facilitate timely treatment decisions.

Second, we implement the classification task within a FL framework to ensure the feasibility of deploying such a model in real-world clinical environments. In this decentralized setting, each participating institution independently trains a local model on its own private CT image datasets. As a result, instead of sharing sensitive patient data, only model parameters or gradients are transmitted to a centralized and trusted server, which aggregates them to construct a globally shared model. This strategy not only preserves data privacy but also enables collaboration across institutions that may have heterogeneous data characteristics, such as variations in imaging protocols, equipment, and patient demographics. Moreover, it facilitates cross-dataset implementation by allowing models to learn from distributed data sources without requiring direct access. Our proposal ensures compliance with data protection regulations by framing kidney stone detection as a federated task.

The next subsections further describe the implementation of our scheme, including the modules that implement it.

A. Image-based Detection of Kidney Stones

Our proposed model frames kidney stone detection as an image-based classification approach for leveraging a DNN architecture. The system is trained on a dataset of CT scans acquired from multiple patients, enabling the model to learn adequate representations of kidney stone characteristics. Through this process, the model is able to automatically extract and distinguish relevant visual features associated with the presence of kidney stones, facilitating accurate classification. Model training is conducted in a supervised manner, where a classification loss function is minimized using backpropagation and stochastic gradient descent. This optimization process allows the model to iteratively refine its parameters, enhancing its ability to generalize across varying image conditions and patient profiles. Once trained, the DNN can effectively analyze previously unseen CT images and infer the presence of kidney stones based on the learned feature space.

B. FL for Kidney Stone Detection

To enhance both the generalization capabilities and achieve privacy-preserving implementation of our image-based kidney stone detection model, we frame it as an FL task. The goal is to enable multiple institutions to collaboratively train the model without sharing raw patient data, thus preserving data privacy and adhering to regulatory requirements. Each institution independently trains a local version of the model on its own CT image dataset, and only model updates are exchanged with a central server for aggregation (Fig. 1). This setup allows the global model to learn from diverse data distributions, achieving cross-dataset performance.

The model training procedure operates following a traditional FL scheme. Given a kidney stone detection function $f(x) : x \rightarrow y$ that outputs the identified label y given an input image x (see Section IV-A). The FL scheme finds an aggregated model w_G, such that w_G is built based on the training data $\{\mathcal{D}_1, \mathcal{D}_2, ..., \mathcal{D}_N\}$, where $\mathcal{D}_i$ denotes the private training dataset from institution I_i, and N represents the number of institutions (such as hospitals).

To achieve such a goal, our scheme is implemented through a four-phase process, namely *Initialization*, *Distribution*, *Local Training*, and *Aggregation*.

1) *Initialization*. The central server initializes a global classification model w_G^t, where t denotes the execution round;

2) *Distribution*. At every communication round t, such that $0 \leq t \leq T$, where T denotes the upper limit of communication rounds, the central server distributes the global model w_G^t to a selected number m of institutions such that $0 \leq m \leq N$, where N denotes the total number of participant institutions;
3) *Local Training*. Each previously selected institution I_i conducts the local training of the received global model w_G^t based on their private training data $\mathcal{D}_i$, compounding a local model w_i^t;
4) *Aggregation*. The central server collects the local models from each selected institution to conduct the model aggregation task that compounds a new global model w_G^{t+1}. The model aggregation task is a function that aggregates a series of models into a single counterpart, e.g., through the Average Federated Learning (FedAVG) function, implemented based on the following equation:

$$w_G^{t+1} = \frac{\sum_{i=0}^{m} w_i^t}{m} \tag{1}$$

where m denotes the number of selected peers. Therefore, the FedAVG builds the new global model w_G^{t+1} by computing the average of each local model w_i^t weights. Finally, if the execution round $t+1$ reaches the upper limit of rounds T, the training is terminated; otherwise, a new *Distribution* phase occurs.

As a result, the FL-based kidney stone detection procedure enables the development of an ML model that is both cross-dataset and privacy-preserving. This is accomplished by training the model locally on each institution's dataset, capturing diverse data distributions, while only sharing model parameters for aggregation. In doing so, the approach maintains data privacy and leverages the heterogeneity of multiple datasets to enhance generalization across clinical environments.

V. Evaluation

Our conducted experiments aim at answering the following Research Questions (RQs):

- ***RQ1*** *- How well do traditional DNN-based approaches perform in cross-dataset kidney stone classification?*
- ***RQ2*** *- Can our proposed scheme maintain high performance across different datasets?*

The next subsections further evaluate the performance of our scheme, including the model-building procedure.

A. Model Building

The proposed model was evaluated considering a binary classification task between *having stone* and *not having stone* classes. To achieve such a goal, the system receives as input the individual's associated cross-sectional CT image and applies the previously trained DNN model (see Section IV-A). We consider a scenario with 2 institutions, each with an associated private dataset as follows:

- ***Institution 1***. Holds a dataset with $1,799$ cross-sectional CT images extracted from 433 subjects, as available in K. Yildirim [5];
- ***Institution 2***. Holds a dataset with $1,027$ cross-sectional CT images extracted from unique individuals, as available in M. N. Islam [21];

Figure 2 overviews a sample of each institution's cross-sectional CT images. We split the original dataset from each institution into *training*, *testing*, and *validation* datasets, each composed by 40%, 30%, and 30% respectively of each behavior. This split is performed considering the distribution for each class to preserve the distribution of kidney stones across all subsets. The *training* set is used to optimize the model parameters, the *validation* set guides hyperparameter tuning, and the *testing* set is reserved exclusively for evaluating the final model's performance.

We implemented our proposal prototype (see Fig. 1) on top of Flower API $v.1.18.0$ [22]. Our proposed image-based detection of kidney stones was implemented through an XResNet-50 DNN architecture, following a similar procedure from K. Yildirim [5]. The architecture was trained for 100 epochs with binary cross entropy as loss function, and its learning rate was set empirically according to the resulting loss and a momentum weight of 0.9. The architecture was implemented through $pytorch$ API $v.2.7.0$ on top of Flower.

For the evaluation of our proposed FL scheme, we considered the evaluation with two institutions (see Fig. 1). Here, each institution is implemented as a separate peer with its corresponding dataset. Each peer is equipped with a 8-core Intel $i7$ CPU, 32 GB of memory, and an Nvidia Tesla T4 GPU running on Ubuntu Linux 22.04. The centralized server also executes the same previously described hardware, but in a separate machine. The upload of local models and the download of their aggregated counterparts are conducted through a network running on a Gigabit interface.

We evaluate the performance of our scheme using the following classification performance metrics:

- *True Positive (TP)*: number of *kidney stone* images correctly classified as *kidney stone*.
- *True Negative (TN)*: number of *normal* samples correctly classified as *normal*.
- *False Positive (FP)*: number of *normal* samples incorrectly classified as *kidney stone*.
- *False Negative (FN)*: number of *kidney stone* images incorrectly classified as *normal*.

Further, we measure the F-Measure according to the harmonic mean of precision and recall values while considering *kidney stone* samples as positive and *normal* samples as negative, as shown in Eq. 4.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

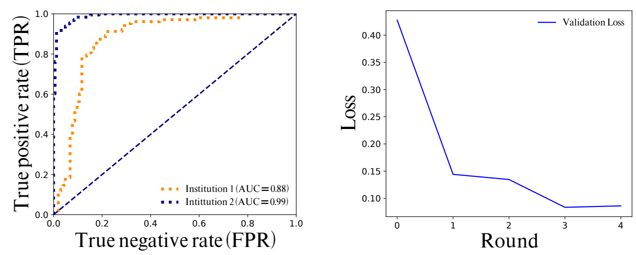
$$F\text{-}Measure = 2 \times \frac{Precision \cdot Recall}{Precision + Recall} \tag{4}$$

(a) Kidney Stone (Institution 1) **(b)** Kidney Stone (Institution 2) **(c)** Normal (Institution 1) **(d)** Normal (Institution 2)

Fig. 2: Sample cross-sectional CT images from each institution. The system is designed as a binary classification task.

TABLE I: Accuracy performance of selected kidney stone detection techniques. Our proposed model achieves high accuracies on a cross-dataset setting.

Scheme	Training Env.	Evaluation Environment					
		Institution 1			Institution 2		
		TP	TN	F1	TP	TN	F1
Trad.	Inst. 1	0.98	0.95	0.96	0.56	0.78	0.70
	Inst. 2	0.46	00.0	0.62	1.00	1.00	1.00
Ours	Federated	0.96	0.93	0.94	0.80	0.85	0.83



(a) ROC curve **(b)** Loss over epochs

Fig. 3: ROC curve and loss of our proposed FL strategy for kidney stone detection.

B. Kidney Stone Detection

Our first experiment aims at answering RQ1, and investigates the classification performance of traditional (centralized) image-based detection of kidney stone images. To achieve such a goal, we train the DNN model using a centralized dataset of CT images collected from a single institution. The resulting model is evaluated on in- and out-of-domain datasets to assess its generalization ability beyond the training distribution. This experiment establishes a performance baseline and highlights the challenges of deploying traditional DNN-based approaches in heterogeneous, real-world clinical environments.

Table I shows the accuracy performance of the traditional centralized approach according to the training environment $vs.$ the evaluation environment. In summary, the model can achieve high accuracy when trained and evaluated within the same institutional dataset. For example, the traditional model trained on Institution 1 provides an F1-score of 0.96 when evaluated on the same environment, demonstrating its effectiveness under consistent data conditions. However, when applied to CT images from a different institution, the F1-score drops to 0.70, indicating a substantial degradation in classification performance. This performance gap indicates that the model cannot operate effectively in cross-dataset scenarios. This limitation arises from domain shifts introduced by variations in imaging protocols, equipment settings, and patient demographics across different institutions. This leads to significant changes in the distribution and visual characteristics of the CT images.

To answer RQ2, we investigate the accuracy performance of our model when operating in a cross-dataset setting. To achieve such a goal, we build the model under an FL setup where each participating institution retains its associated dataset locally, simulating a realistic distributed medical environment. In practice, we consider a scenario with two peers, where one peer builds a model using the dataset from Institution 1, and the other peer uses the dataset from Institution 2 (see Section V-A). The local models trained at each institution are then aggregated using the FedAvg algorithm to produce a global model (see Section IV-B). This evaluation aims to assess whether our proposed scheme can effectively learn from heterogeneous data sources while preserving privacy and still achieving high classification accuracy across all participating datasets, without data centralization.

Table I shows the accuracy performance of our model when operating following a federated approach. Our proposed scheme successfully enables a cross-dataset implementation with high accuracy, demonstrating its effectiveness in real-world scenarios where data privacy must be preserved. In practice, the aggregated model achieves high classification performance across both datasets, with an F1-score of 0.94 in Institution 1 and 0.83 in Institution 2. Compared to the traditional centralized approach, our scheme achieves a higher average F1-Score of 0.89, whereas the traditional models reach only 0.83 and 0.81 when trained on datasets from Institution 1 and Institution 2, respectively. Figure 3a shows the obtained Receiver Operating Characteristic (ROC) curves of our proposed scheme for each institution. The model achieves an Area Under the Curve (AUC) of 0.88 for Institution 1 and 0.99 for Institution 2, demonstrating strong discriminative performance across both datasets. As a result, the federated setup allows participating institutions to collaboratively build a reliable detection model without exchanging sensitive patient data. These results highlight the potential of our FL-based approach to support accurate kidney stone detection across heterogeneous medical environments while meeting privacy constraints.

VI. Conclusion

Kidney stone detection from Computed Tomography (CT) images is a challenging task in clinical diagnosis, where building effective models often requires access to large and diverse datasets that are typically siloed across institutions. Unfortunately, sharing such medical data across institutions is often not feasible due to strict privacy regulations and patient confidentiality concerns. In this work, we propose a Deep Neural Network (DNN)-based scheme for image-based kidney stone detection using CT scans, implemented as a twofold process. First, we design it to automatically extract relevant visual patterns to classify the presence of kidney stones. Second, to preserve privacy and enhance generalization, we frame the model's implementation as a Federated Learning (FL) task, allowing institutions to collaboratively train without sharing raw data. The FL training process is conducted locally at each institution and then aggregated to form a global model. Experimental results demonstrate that our proposed model achieves high classification accuracy within the same training environment, with an F1-score of up to 0.94. In addition, in cross-dataset evaluations, our approach outperforms traditional centralized baselines, showing significantly lower performance degradation while preserving patient privacy.

Future research may enhance the proposed Federated Learning framework by incorporating a multiview strategy that leverages complementary information from axial, coronal, and sagittal planes for improved model generalization and diagnostic performance.

Acknowledgment

This work was partially sponsored by the Brazilian National Council for Scientific and Technological Development (CNPq), grants nº 304990/2021-3, 407879/2023-4, 302937/2023-4, 406030/2023-5, and 306878/2022-4, Araucária Foundation, the Secretariat of Science, Technology and Higher Education (SETI) and the Coordination for the Improvement of Higher Education Personnel (CAPES).

References

[1] A. Garg and V. Mago, "Role of machine learning in medical research: A survey," *Comput. Sci. Rev.*, vol. 40, p. 100370, May 2021.

[2] K. K. Patro, J. P. Allam, B. C. Neelapu, R. Tadeusiewicz, U. R. Acharya, M. Hammad, O. Yildirim, and P. Pławiak, "Application of kronecker convolutions in deep learning technique for automated detection of kidney stones with coronal CT images," *Inf. Sci. (Ny)*, vol. 640, no. 119005, p. 119005, Sep. 2023.

[3] S. Gerke and D. Rezaeikhonakdar, "Privacy aspects of direct-to-consumer artificial intelligence/machine learning health apps," *Intell. Based Med.*, vol. 6, no. 100061, p. 100061, 2022.

[4] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[5] K. Yildirim, P. G. Bozdag, M. Talo, O. Yildirim, M. Karabatak, and U. R. Acharya, "Deep learning model for automated kidney stone detection using coronal ct images," *Computers in Biology and Medicine*, vol. 135, p. 104569, 2021.

[6] J. Geremias, E. K. Viegas, A. O. Santin, A. Britto, and P. Horchulhack, "Towards a reliable hierarchical android malware detection through image-based cnn," in *2023 IEEE 20th Consumer Communications amp; Networking Conference (CCNC)*. IEEE, Jan. 2023, p. 242–247.

[7] H. Guan, P.-T. Yap, A. Bozoki, and M. Liu, "Federated learning for medical image analysis: A survey," *Pattern Recognit.*, vol. 151, no. 110424, p. 110424, Jul. 2024.

[8] R. R. dos Santos, E. K. Viegas, A. O. Santin, and P. Tedeschi, "Federated learning for reliable model updates in network-based intrusion detection," *Computers amp; Security*, vol. 133, p. 103413, Oct. 2023.

[9] M. J. Sheller, B. Edwards, G. A. Reina, J. Martin, S. Pati, A. Kotrotsou, and S. Bakas, "Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data," *Scientific Reports*, vol. 10, no. 1, p. 12598, 2020.

[10] M. Y. Lu, D. F. K. Williamson, T. Y. Chen, R. J. Chen, M. Barbieri, and F. Mahmood, "Federated learning for computational pathology on gigapixel whole slide images," *Medical Image Analysis*, vol. 76, p. 102298, 2022.

[11] X. Li, Y. Gu, N. C. Dvornek, P. Ventola, and J. S. Duncan, "Multi-site fmri analysis using privacy-preserving federated learning and domain adaptation: Abide results," *Medical Image Analysis*, vol. 65, p. 101765, 2020.

[12] S. Sundaramoorthy and K. Jayachandru, "Designing of enhanced deep neural network model for analysis and identification of kidney stone, cyst, and tumour," *SN Comput. Sci.*, vol. 4, no. 5, pp. 1–7, Jun. 2023.

[13] V. M. de Oliveira, H. M. de Oliveira, G. M. Santos, J. Geremias, and E. K. Viegas, "A big data framework for scalable and cross-dataset capable machine learning in network intrusion detection systems," *IEEE Access*, vol. 13, p. 129419–129431, 2025.

[14] J. Porkodi, M. Thenmozhi, and R. Priyadharshini, "Detection and classification of kidney stone in ct images," *International Journal of Scientific Research in Science and Technology (IJSRST)*, vol. 5, no. 5, pp. 184–188, March-April 2020, print ISSN: 2395-6011.

[15] A. Pimpalkar, D. K. J. B. Saini, N. Shelke, A. Balodi, G. Rapate, and M. Tolani, "Fine-tuned deep learning models for early detection and classification of kidney conditions in ct imaging," *Scientific Reports*, vol. 15, p. 10741, 2025.

[16] K. Sharma, Z. Uddin, A. Wadala, and D. Gupta, "Hybrid deep learning framework for classification of kidney ct images: Diagnosis of stones, cysts, and tumors," *arXiv preprint arXiv:2502.04367*, 2025.

[17] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*. PMLR, 2017, pp. 1273–1282.

[18] A. Caglayan, M. O. Horsanali, K. Kocadurdu, E. Ismailoglu, and S. Guneyli, "Deep learning model-assisted detection of kidney stones on computed tomography," *International Brazilian Journal of Urology*, vol. 48, no. 5, pp. 830–839, 2022.

[19] V. Vekaria, R. Gandhi, B. Chavarkar, H. Shah, C. Bhadane, and P. Chaudhari, "Identification of kidney disorders in decentralized healthcare systems through federated transfer learning," in *Procedia Computer Science*, vol. 233. Elsevier, 2024, pp. 998–1010.

[20] A. G. Filho, E. K. Viegas, A. O. Santin, and J. Geremias, "A dynamic network intrusion detection model for infrastructure as code deployed environments," *Journal of Network and Systems Management*, vol. 33, no. 4, Jun. 2025.

[21] M. N. Islam, M. Hasan, M. Hossain, M. Alam, G. Rabiul, M. Z. Uddin, and A. Soylu, "Vision transformer and explainable transfer learning models for auto detection of kidney cyst, stone and tumor from ct-radiography," *Sci. Reports*, vol. 12, no. 1, pp. 1–4, July 2022.

[22] D. J. Beutel, T. Topal, A. Mathur, X. Qiu, J. Fernandez-Marques, Y. Gao, L. Sani, H. L. Kwing, T. Parcollet, P. P. d. Gusmão, and N. D. Lane, "Flower: A friendly federated learning research framework," *arXiv preprint arXiv:2007.14390*, 2020.