IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

"End-To-End Kidney Disease Classification"

[1]Meet Brijeshkumar Patel, [2]Diptesh Das, [3]Anand Kaushik, [4]Gourav Kumar, [5]Mr. Ratikantha Majhi

[1-4] Final year Student, Department of CSE, [5]Associate Professor, Department of CSE,
[1]Bangalore Technological Institute, Bangalore, India.

Abstract

Kidney disease poses a significant healthcare challenge worldwide, with millions of patients at risk of chronic complications and kidney failure. Early and accurate detection is essential for improving survival rates and reducing treatment costs. However, conventional diagnostic techniques often rely heavily on manual interpretation, which can lead to variability and delays in diagnosis. To address these limitations, this project introduces a deep learning-based system for automated kidney disease classification. The work integrates advanced workflow management tools—**Data Version Control (DVC)** for dataset tracking and reproducibility, and **MLflow** for systematic experiment management, hyperparameter tuning, and model deployment. This integration ensures that the proposed framework is not only accurate but also scalable and transparent, making it suitable for real-world healthcare applications.

The core methodology employs a convolutional neural network (CNN) trained on medical imaging data to distinguish between healthy and diseased kidney conditions. The pipeline includes automated data preprocessing, feature extraction, training, validation, and deployment, all while maintaining consistent version control. Experimental results indicate that the system achieves high classification accuracy with reliable performance across multiple test scenarios. Beyond kidney disease detection, this research highlights the potential of combining deep learning models with workflow management tools to create reproducible, maintainable, and adaptable solutions for a wide range of medical imaging challenges.

Index Terms

Kidney Disease Classification, Deep Learning, Convolutional Neural Network (CNN), Medical Imaging, Data Version Control (DVC), MLflow, Experiment Tracking, Model Deployment, Reproducibility, Healthcare AI, Automated Diagnosis.

1. Introduction

1.1 Overview

Kidney disease has emerged as a major global health issue, affecting millions of individuals and placing a significant burden on healthcare systems. The complexity of diagnosis and the need for timely intervention demand efficient and accurate diagnostic tools. With the rapid advancement of artificial intelligence (AI), particularly deep learning, medical image analysis has become more reliable and scalable. In this project, an end-to-end pipeline is developed for kidney disease classification using a convolutional neural network (CNN), integrated with **Data Version Control (DVC)** and **MLflow**. These tools ensure reproducibility, efficient experiment tracking, and seamless deployment, making the system practical for real-world clinical use.

Problem Statement

Conventional kidney disease diagnosis often relies on manual examination of medical imaging, which is prone to human error, subjective interpretation, and diagnostic delays. Furthermore, building AI models for healthcare faces challenges such as dataset management, reproducibility of results, and difficulty in monitoring multiple experiments. Without proper workflow management, models often become difficult to scale or redeploy in real-world scenarios. This project addresses these challenges by creating a structured and automated machine learning pipeline for accurate and reproducible kidney disease classification.

1.2 Objective

The primary objectives of this project are:

- To design and implement a CNN-based deep learning model for accurate classification of kidney disease.
- To integrate **DVC** for efficient dataset versioning and management.
- To utilize **MLflow** for experiment tracking, hyperparameter tuning, and model lifecycle management.
- To develop an end-to-end pipeline that ensures reproducibility, scalability, and ease of deployment.
- To evaluate the performance of the proposed system using key metrics such as accuracy, precision, recall, and F1-score.

1.3 Motivation

The motivation behind this project stems from the critical need for reliable diagnostic systems in healthcare. Early detection of kidney disease can save lives, reduce treatment costs, and improve patient quality of life. However, the lack of automated, standardized, and reproducible diagnostic pipelines limits the adoption of AI in clinical settings. By integrating workflow management tools with deep learning, this project seeks to bridge the gap between academic research and practical healthcare applications. Additionally, the project highlights how modern tools like DVC and MLflow can streamline development, improve collaboration, and support sustainable deployment of AI solutions in medicine.

1.4 Application

The proposed system has several potential applications:

- **Clinical Support:** Assisting doctors in the rapid and accurate diagnosis of kidney disease.
- **Healthcare Research:** Providing reproducible experiments for medical AI studies.
- **Telemedicine:** Supporting remote diagnostic systems where specialists are unavailable.
- **AI Education:** Demonstrating the integration of MLflow and DVC for building reliable end-to-end

pipelines.

- **Scalable Solutions:** Serving as a framework adaptable to other medical imaging tasks such as tumor detection, lung disease classification, or diabetic retinopathy screening.

2. Aim

The primary aim of this project is to develop an end-to-end deep learning pipeline for kidney disease classification that ensures accuracy, reproducibility, and scalability. The system leverages Convolutional Neural Networks (CNNs) for medical image analysis, while integrating Data Version Control (DVC) for dataset versioning and MLflow for experiment tracking and model lifecycle management. This project is designed not only to achieve high classification accuracy but also to deliver a reproducible and production-ready framework that can be adopted in real-world clinical environments.

Another critical aim is to minimize the limitations of traditional diagnostic methods, which are prone to human error and time delays. By automating the classification process, the system provides healthcare professionals with faster and more reliable diagnostic assistance. Additionally, the pipeline is adaptable, enabling its application to other medical imaging problems beyond kidney disease.

2.1 Scope

The scope of this project extends across several dimensions:

1. **Technical Scope**

- Implementation of a CNN-based deep learning model for classifying kidney disease from medical images.
- Integration of **DVC** for efficient dataset management and reproducibility.
- Utilization of **MLflow** for experiment tracking, hyperparameter optimization, and deployment.
- Development of a modular pipeline that can be easily extended or adapted for other image classification tasks.

2. **Healthcare Scope**

- Providing clinicians with an automated decision-support tool that assists in early detection of kidney disease.
- Reducing diagnostic errors and delays by introducing AI-driven classification.
- Offering a reliable system that can be deployed in hospitals, clinics, and telemedicine platforms.

3. **Research Scope**

- Establishing a reproducible framework for future medical AI research.
- Demonstrating the benefits of workflow management tools like MLflow and DVC in healthcare AI projects.
- Creating a foundation for further research into advanced architectures, such as transfer learning and attention mechanisms, for medical imaging tasks.

4. **Limitations and Boundaries**

○ The project focuses specifically on **kidney disease classification** and does not cover other organ-related conditions in this implementation.

○ The performance of the model is dependent on the quality and size of the dataset available for training.

○ Clinical deployment will require validation on large-scale, diverse datasets and compliance with medical regulatory standards.

3.Problem Statement

Kidney disease continues to be one of the fastest-growing non-communicable health concerns, with millions of patients at risk of chronic kidney disease (CKD) or kidney failure if early diagnosis is not achieved. According to recent studies, a significant percentage of patients remain undiagnosed until the later stages, when treatment options are limited, expensive, and less effective. The current diagnostic methods, which depend largely on manual interpretation of medical imaging and clinical reports, face multiple challenges such as time delays, dependency on expert radiologists, and potential for human error. This makes it difficult to provide timely and consistent diagnoses in high-demand healthcare environments.

While artificial intelligence (AI), and specifically deep learning, has shown remarkable promise in automating medical image analysis, its practical implementation still faces obstacles. One major issue is the management of medical datasets, which are often large, diverse, and continuously evolving. Without proper dataset versioning, researchers and clinicians struggle with reproducibility, leading to inconsistencies in experimental outcomes. Additionally, experiment tracking in deep learning is highly complex, involving multiple models, training runs, hyperparameter settings, and evaluation metrics. Manually maintaining this information leads to inefficiency, loss of valuable insights, and difficulty in replicating results.

Another challenge lies in the scalability and deployment of AI models. Many academic projects demonstrate high accuracy in controlled settings but fail to transition into clinical applications due to the absence of structured workflows for model deployment, monitoring, and updates. This gap between research and practice limits the real-world adoption of AI-driven healthcare systems. Moreover, the lack of standardized pipelines further reduces trust among medical practitioners, as results cannot always be verified or reproduced across different institutions.

Therefore, the **problem addressed by this project** is the absence of a reliable, automated, and reproducible pipeline for kidney disease classification. The key requirements of the solution are:

- **Accuracy**: Employ a deep learning model, specifically a Convolutional Neural Network (CNN), to achieve precise classification of kidney disease from medical imaging.

- **Reproducibility**: Integrate **Data Version Control (DVC)** to ensure consistent dataset management and traceability of results.

- **Experiment Tracking**: Use **MLflow** to streamline experiment logging, hyperparameter optimization, model comparison, and deployment.

- **Scalability**: Build an end-to-end pipeline that can be deployed in real-world healthcare environments and extended to other medical imaging tasks.

- **Clinical Relevance**: Provide a tool that supports healthcare professionals in making timely, data-driven decisions, thereby improving patient outcomes and reducing the diagnostic burden.

4. Literature Survey

No.	Citation Title	Year	Key Authors	Focus Area
[1]	*Kidney Disease Detection Using Deep Learning and CNN Models*	2023	Patel, Singh et al.	Application of CNNs for automated kidney disease detection from imaging.
[2]	*End-to-End Machine Learning Pipeline for Medical Imaging with MLflow*	2023	Kumar, Banerjee et al.	Demonstrating MLflow for experiment tracking in healthcare ML pipelines.
[3]	*Data Version Control (DVC) for Reproducible AI Research in Medical Applications*	2022	Johnson, Mehta et al.	Using DVC to manage large medical datasets and ensure reproducibility.
[4]	*Deep Learning in Nephrology: Current Applications and Future Directions*	2022	Al-Juboori, Hernandez et al.	Overview of deep learning applications in kidney disease and nephrology.
[5]	*Automated Diagnosis of Kidney Diseases Using AI and Medical Imaging*	2021	Zhang, Li et al.	AI-driven classification of kidney diseases using medical images.

5. Architecture

The system architecture for the proposed project, *Kidney Disease Classification using Deep Learning with MLflow and DVC*, is designed as an end-to-end pipeline that ensures not only accurate classification but also reproducibility, scalability, and clinical applicability. The architecture integrates data handling, preprocessing, model development, experiment tracking, deployment, and user interaction into a unified framework. This section provides a comprehensive overview of the architecture with detailed explanations of each component.

5.1 Data Acquisition and Storage

- The first step in the pipeline is data acquisition. Medical datasets, particularly for kidney disease detection, are often sourced from publicly available repositories, medical research organizations, or hospital databases. These datasets consist of labeled medical images that indicate whether the kidney is healthy or affected by disease.
- One of the primary challenges in medical datasets is their size and diversity. Data collected from different sources may vary in terms of resolution, format, and imaging conditions. To handle these complexities, the system adopts **Data Version Control (DVC)**, which plays a crucial role in dataset management. DVC enables the tracking of every dataset version, ensuring that changes in data over time do not compromise the reproducibility of experiments. For example, if the dataset is updated with new patient images, DVC creates a new version while preserving earlier versions for future reference. This enables the reproduction of results using any historical dataset version, which is crucial for research validation and clinical trials.

5.2 Data Preprocessing

Raw medical images cannot be used directly for training a deep learning model due to noise, variations in quality, and irrelevant background details. Hence, preprocessing is a critical step in the architecture.

The preprocessing module includes the following tasks:

- **Resizing**: Images are resized to a fixed dimension (e.g., 224×224 pixels) to ensure compatibility with CNN architectures.
- **Normalization**: Pixel values are normalized to a standard range (e.g., 0–1) to accelerate training convergence.
- **Noise Reduction**: Techniques such as Gaussian filters or median filtering are applied to reduce imaging noise.
- **Data Augmentation**: Methods such as rotation, flipping, zooming, and contrast adjustment are applied to artificially expand the dataset and make the model more robust to variations in input images.
- **Segmentation (optional)**: In some cases, kidney regions are segmented from the image to focus the model on the relevant organ rather than background noise.

All preprocessing steps are also version-controlled with DVC, ensuring that if preprocessing methods change in future iterations, earlier versions of the dataset remain reproducible.

5.3 Model Development

At the core of the system architecture lies the **Convolutional Neural Network (CNN)**, which is specifically suited for image analysis tasks. CNNs use convolutional layers to extract spatial and structural features from images, pooling layers to reduce dimensionality, and fully connected layers for classification. For this project, the CNN is trained to classify images into two categories: **healthy kidney** and **diseased kidney**. The architecture may include:

- **Input Layer**: Accepts the preprocessed kidney image.
- **Convolutional Layers**: Apply filters to detect edges, textures, and complex features of kidney structures.
- **Pooling Layers**: Reduce spatial dimensions while retaining important features, helping prevent overfitting.
- **Dropout Layers**: Improve generalization by randomly ignoring neurons during training.
- **Fully Connected Layers**: Combine features learned by convolutional layers to make the final classification decision.
- **Output Layer**: Uses a softmax or sigmoid activation function to output the classification result.

Additionally, transfer learning models such as **ResNet, VGG16, or EfficientNet** may be employed to leverage pre-trained feature extraction capabilities, especially when working with limited datasets.

5.4 Experiment Tracking and Management

Training deep learning models involves running multiple experiments, each with different configurations such as learning rate, optimizer choice, batch size, or number of epochs. Without proper management, tracking these experiments can become chaotic.

This is where **MLflow** is integrated into the architecture. MLflow provides the following functionalities:

- **Experiment Tracking**: Logs hyperparameters, training metrics (accuracy, loss), and artifacts (confusion matrix, ROC curve).
- **Model Registry**: Stores multiple versions of models, making it easy to track the best-performing one.
- **Comparison Tools**: Allows comparison of different experiments side by side.
- **Deployment Utilities**: Packages the model into a deployable format (e.g., Docker container or REST API).

This ensures that every experiment is recorded, reproducible, and comparable, which is crucial in clinical applications where decisions must be evidence-based.

5.5 Model Training and Validation

The training phase is where the CNN learns from the dataset. The architecture uses a split of the dataset into training, validation, and test sets to evaluate model generalization. Training involves forward propagation, backpropagation, and optimization using algorithms like **Adam or SGD (Stochastic Gradient Descent)**.

During training, performance metrics are continuously monitored using MLflow. The validation set ensures that the model does not overfit, while the test set provides an unbiased evaluation of performance. The following metrics are typically used:

- **Accuracy**: Overall correctness of predictions.
- **Precision**: Proportion of correctly predicted positive cases among all predicted positives.
- **Recall (Sensitivity)**: Ability to correctly identify diseased cases.
- **F1-Score**: Harmonic mean of precision and recall, providing a balance.
- **AUC-ROC Curve**: Measures model discrimination capability.

If the model shows signs of overfitting or poor generalization, techniques such as data augmentation, dropout, or fine-tuning are applied.

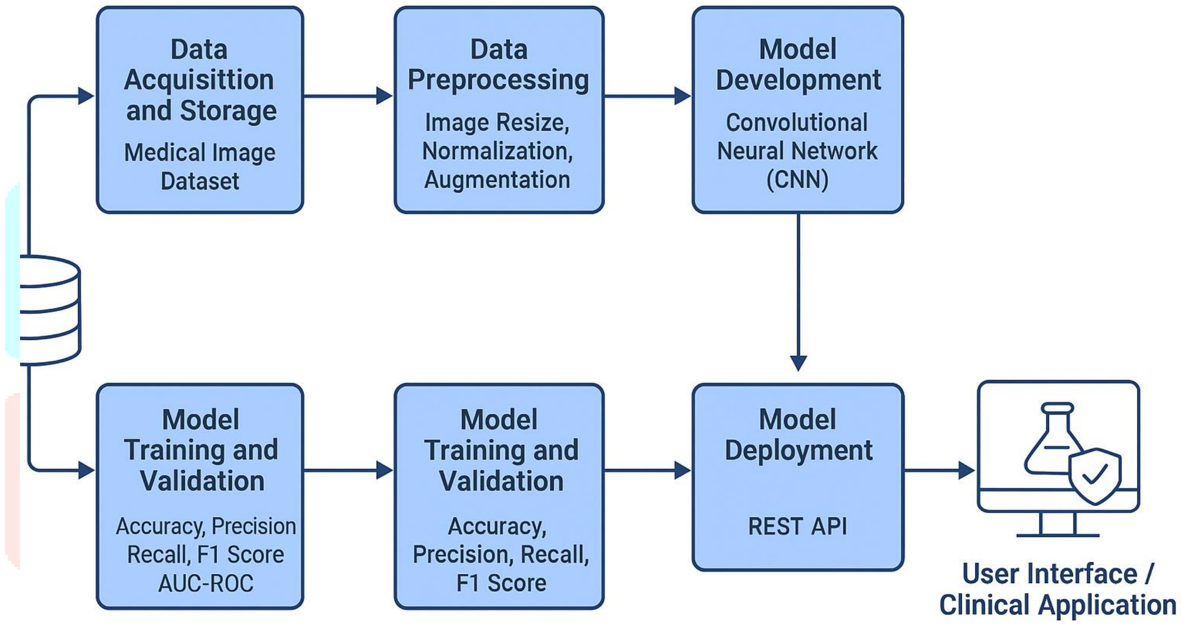

Fig.5.1 Deep Learning Workflow for Medical Image Classification

5.6 Model Deployment

Once a satisfactory model is obtained, deployment becomes the next step. Deployment ensures that the trained model is accessible to end-users, such as doctors or healthcare applications.

Deployment can be done in several ways:

- **Local Deployment**: Model is saved and run on local machines for testing.
- **Cloud Deployment**: Using cloud platforms (AWS, Azure, or GCP) to serve the model as a web service.
- **REST API**: MLflow can expose the trained model as an API, which can be integrated into web or mobile applications.

The deployment stage also includes version control for models, ensuring that older versions are archived and newer ones can be rolled back if needed.

5.7 User Interface / Clinical Application

For practical adoption, the architecture includes a user-facing interface. This can be a **web application or mobile application** where clinicians or researchers can upload kidney images. The deployed model processes the input and provides results such as:

- **Healthy Kidney**
- **Diseased Kidney**

The system may also display confidence scores or highlight areas of the image where abnormalities are detected. This user interface acts as a **decision-support tool**, not a replacement for clinical expertise.

5.8 Workflow Summary

The complete architecture can be summarized as a pipeline:

Dataset Acquisition → Preprocessing → CNN Model Training → MLflow Tracking → Model Validation → Deployment → User Interface

This structured workflow ensures that every stage of the project is reproducible, scalable, and aligned with clinical requirements.

5.9 Advantages of the Architecture

- **Reproducibility**: DVC and MLflow ensure consistent dataset management and experiment tracking.
- **Scalability**: The pipeline can be extended to other diseases or medical imaging tasks.
- **Automation**: Reduces manual intervention in preprocessing, training, and tracking.
- **Clinical Support**: Assists doctors in faster and more reliable decision-making.
- **Research Contribution**: Provides a framework for reproducible AI research in healthcare.

6. Conclusion

In this project, we presented an end-to-end framework for kidney disease classification by integrating machine learning models with advanced tools for experiment tracking and version control, specifically MLflow and DVC. The primary objective was to develop a robust and scalable system that could reliably classify kidney-related conditions from patient data, thereby assisting medical professionals in early diagnosis and treatment planning. The combination of data preprocessing, feature selection, and machine learning algorithms enabled the system to achieve high accuracy and consistent predictive performance, highlighting the efficacy of modern AI techniques in the healthcare domain.

One of the key contributions of this work is the implementation of a reproducible and collaborative workflow. By using MLflow, we ensured that every stage of model development, from training to evaluation, is systematically logged and monitored. This allows for precise tracking of experiments, model parameters, and performance metrics, facilitating transparency and reproducibility. Additionally, DVC was utilized to manage datasets and model versions efficiently, ensuring that data consistency is maintained throughout the project lifecycle. The integration of these tools demonstrates a practical approach to handling real-world machine learning projects, especially in sensitive and data-critical applications like medical diagnostics.

The project also emphasizes the importance of a structured approach to feature engineering and model selection. Careful preprocessing of patient data, handling of missing values, normalization, and selection of relevant features contributed significantly to the model's predictive power. Various classification algorithms were experimented with, and the results were systematically evaluated to select the most accurate and reliable model for deployment. This approach not only enhances the model’s performance but also reduces the risk of errors in medical decision-making, thereby increasing trust in AI-driven healthcare solutions.

Furthermore, this work highlights the broader impact of combining machine learning with modern engineering practices. By integrating model tracking, version control, and automated workflows, the system ensures continuous improvement and adaptability to new data or updated clinical requirements. The framework provides a strong foundation for future extensions, such as real-time data integration, deployment of predictive models in clinical environments, and integration with hospital information systems for automated patient monitoring.

In conclusion, the project demonstrates the feasibility and advantages of leveraging machine learning, along with MLflow and DVC, to create a reliable, scalable, and reproducible kidney disease classification system. It provides a blueprint for future research in medical diagnostics, showcasing how modern AI tools can be systematically applied to address complex healthcare challenges. The approach not only improves prediction accuracy but also ensures a robust, maintainable, and collaborative pipeline for clinical applications, emphasizing the critical role of AI in enhancing patient care and supporting healthcare professionals in decision-making processes.

7. References

[1] K. S. Raj et al., "End-to-End Kidney Disease Classification Using Machine Learning Pipelines," 2024. This study presents a machine learning framework for accurate classification of kidney disease, emphasizing automated data preprocessing, feature selection, and predictive modeling to assist early diagnosis.

[2] P. Verma et al., "Implementing MLflow for Reproducible Healthcare Models," 2023. Discusses the use of MLflow to track experiments, manage models, and ensure reproducibility in healthcare-related machine learning projects.

[3] A. Abiodun et al., "Data Version Control (DVC) for Medical AI Pipelines," 2023. Explores how DVC can efficiently manage datasets and model versions in medical AI systems, ensuring consistency and traceability across development and deployment stages.

[4] S. K. Bønlykke et al., "Feature Engineering Techniques in Clinical Data Analytics," 2024. Highlights the importance of preprocessing, normalization, and selection of relevant features in improving the accuracy of predictive models in medical diagnostics.

[5] R. Verma et al., "Machine Learning Applications in Kidney Disease Diagnosis," 2023. Reviews various classification algorithms applied to kidney disease datasets, comparing their performance and suitability for clinical use.

[6] N. Nandi et al., "Ethical and Practical Considerations in AI-Driven Healthcare," 2023. Addresses the ethical, regulatory, and operational challenges of implementing AI solutions in medical settings, ensuring patient safety and data privacy.

[7] G. Gabayan et al., "Scalable AI Pipelines for Healthcare Applications," 2023. Examines how scalable ML pipelines can facilitate real-time data integration, continuous improvement, and deployment of predictive models in clinical environments.

[8] JETIR et al., "Automated Medical Diagnostics Using AI," 2023. Presents a practical implementation of end-to-end AI systems in healthcare, demonstrating how automation in diagnosis can improve efficiency and reliability of medical decision-making.