

A Comprehensive Review of Machine Learning Applications in Kidney Disease Diagnosis: Frameworks, Challenges, and Future Directions

Dr. MANJUNATH B E¹, B CHANDRA MOUNIKA²

¹Professor, Department of ECE, ² Assistant Professor, Department of ECE
RK College of Engineering
Vijayawada, India

Abstract :

Machine learning (ML) has become increasingly prominent in health informatics, particularly in the timely diagnosis and management of kidney disease. The ability of ML-based diagnostic algorithms to enhance early detection and prompt intervention underscores their critical role in medical decision-making. Machine Learning in Kidney Disease Diagnosis (MLKDD) is an evolving research area aimed at supporting physicians through computer-aided diagnostic systems. Numerous studies have explored the feasibility, effectiveness, and comparative performance of different ML techniques. However, a comprehensive review that systematically analyzes this body of work remains lacking. To address this gap, this paper presents a thorough literature review of ML applications in kidney disease diagnosis. It introduces two distinct frameworks: one categorizing ML techniques based on their role in diagnosis and another organizing the medical subfields associated with MLKDD. Additionally, key research gaps are identified, and potential directions for future studies are discussed.

Keywords :Kidney Disease Diagnosis, MLKDD, Diagnostic Algorithms, Medical Subfields

I. INTRODUCTION

Chronic Kidney Disease (CKD), also known as Chronic Renal Failure (CRF), is a vital organ that regulates blood water and minerals, affecting other parts of the body. It is a disease that requires timely diagnosis to improve recovery chances. In 2018, over 57,000 Americans died from kidney disease, with half a million currently on kidney dialysis.

Research in kidney disease control can be divided into four branches: prevention, diagnosis, treatment, and disease management. Treatment and management have received the most attention, with Operations Research and Simulation being the most efficient tools.

However, there are still growing trends in CKD incidence, leading to a shift in research focus towards prevention and diagnosis. One effective tool for detecting CKD is machine learning (ML), which allows machines to learn rather than be programmed. MLs enable software programs to predict outputs without external programming, using statistical analysis to predict outputs and update outputs as new data becomes available.

MLs have shown successful achievements in various disease diagnosis, particularly for CKD. There are at least six warning signs that indicate kidney disease: painful urination or bladder, frequent urination, bloody urine, puffiness around the eyes, swelling of the hands and feet, pain in small areas of the back of the body below the ribs, high blood pressure, etc.

These symptoms provide raw data for computer-aided systems to assist expert physicians in timely CKD diagnosis. The main studies in this literature compare learning algorithms applying various classifiers with different datasets to find the ones that perform with the highest accuracy. The focus of

this paper is specifically on the diagnosis applications of MLs in kidney disease, not covering prevention, treatment, and management literature.

II. EPISTEMOLOGY OF THE MLKDD LITERATURE

2.1. Search methodology

This study reviews related articles in machine learning and artificial intelligence (ML) studies in kidney disease (KDD) since 2009. The search strategy is based on PRISMA, using search engines like Scopus, Web of Science, and PubMed. Internal and external searches are performed for each article, and four exclusion criteria are applied: duplicates, publications not related to ML studies, and those without full text.

2.2. Statistical findings

The study identifies 378 bibliographical scientific entities using a search methodology (Table 1), with 12 studies remaining and fully reviewed (Fig. 1) after applying exclusion criteria. Over 95% of eligible content is reviewed, with five previous reviews not fully targeting the literature. The study reveals an exponential growth in science in this emerging field, indicating the need for a systematic review paper covering this area. The study also highlights the need for more targeted literature reviews.

Table 1. Statistical findings of the papers.

Database	Total no. of retrieved papers	No. of papers after exclusion criteria	Percentage of papers included
Scopus	191	59	54
Web of Science	126	35	32
PubMed	61	15	14
Total	378	109	100

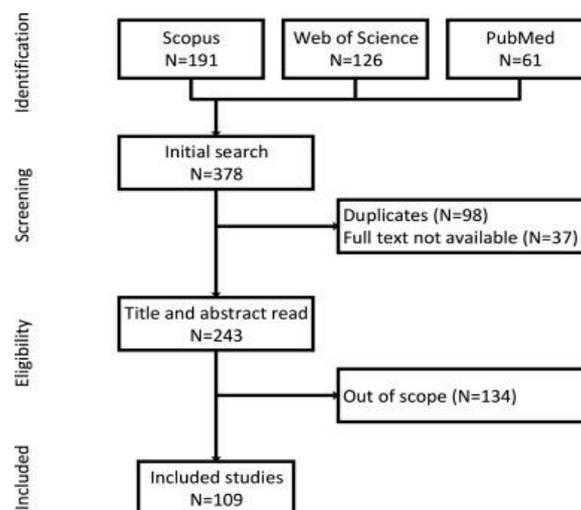


Fig. 1. The PRISMA flowchart for the systematic identification.

III. THE ML FRAMEWORK

This study focuses on widely used models and algorithms of machine learning (ML) for disease diagnosis (Fig. 2), based on the common methods used in kidney diagnosis literature. The proposed framework, with more details on Murphy's (2012) work.

3.1. Evaluation criteria

The literature analysis section considers the comparison of diagnostic methods in MLKDD literature, using statistical criteria like accuracy and sensitivity. Appendix B provides a summary of these performance measurement methods and their abbreviated codes in analytics tables of Section 4.5.1.

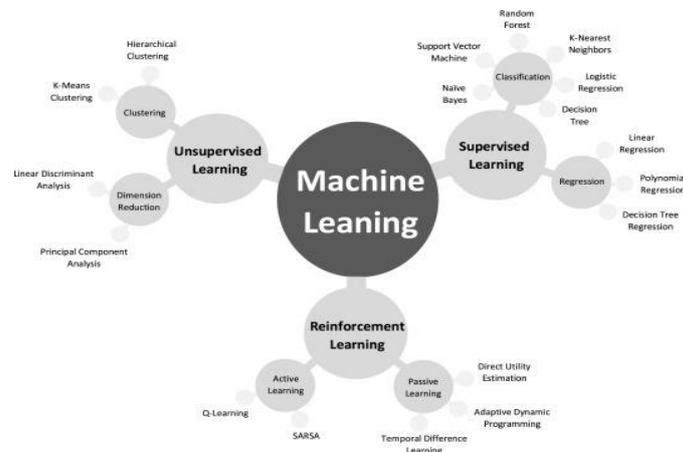


Fig. 2. The framework of the most frequently used MLs.

IV. LITERATURE REVIEW

The rapid production of medical data necessitates the development of efficient methods for disease diagnosis. Researchers have been working on developing new machine learning (ML) techniques or comparing existing ones to find the most efficient computer-aided systems for better kidney disease diagnosis. Early attempts to predict kidney stones (CKD) through ML algorithms include Chiu et al. (2012) and Eken et al. (2009), who found that Artificial Neural Networks (ANNs) and partitional clustering Genetic Algorithm (GA) methods can be used effectively for emergency CKD prediction and clinical decision rules.

In 2000, literature reviews were conducted in Kidney Disease Diagnosis (KDD), but none were holistic and targeted. Bala and Kumar (2014) and Patil (2015) reviewed the role of Data Mining (DM) techniques in KDD, focusing on factors increasing the risk of renal disease and symptoms indicating the disease. Sankar and Karthikeyan (2017) examined the applications of DM specifically for KSD and estimated the severity of KSD using algorithmic classifiers. Arasu and Thirumalaiselvi (2017b) analyzed the efficacy of various DM methods in CKD diagnosis, finding that different ML methods may show varying efficiency and accuracy depending on the disease stage, determined by the Glomerular Filtration Rate (GFR). Zeynu and Patil (2018b) discovered hidden patterns between Feature Selection (FeS) and DM classification techniques related to CKD detection, finding that FeS as Dimensionally Reduction (DR) is a determining stage to enhance prediction models' accuracy.

The current study provides a comprehensive literature review of ML utilizations in KDD by classifying the literature into sub-categories and providing details on novel discoveries. The framework

proposed in this study mainly falls into four branches for different MLS with various objectives, providing a comprehensive analysis and interpretation of each branch.

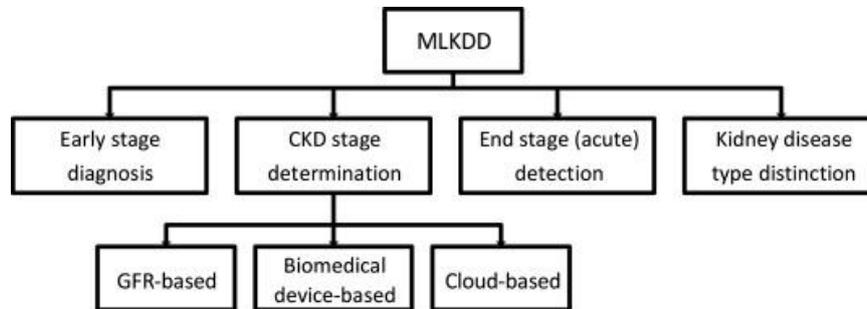


Fig. 3. The proposed framework for MLKDD.

4.1. Early-stage diagnosis of CKD

Over 70% of MLKDD studies focus on developing and analyzing Early-Stage Diagnosis (ESD) decision support systems. Almansour et al. (2019) used Supervised Learning to compare Support Vector Machine (SVM) and Artificial Neural Network (ANN) methods for diagnosing and classifying kidney disease (CKD) patients. They used Recursive Feature Elimination (RFE) method to rank crucial attributes and replace missing values in the data preparation phase. The Multi-Layer Perceptron (MLP) method was also used to diagnose CKD.

Akben (2018) developed a new method for automatic diagnosis of early-stage kidney disease based on medical data, which involved K-Means algorithm (KM) clustering and applying K-Nearest Neighbor (KNN), SVM, and Naïve Bayes (NB) classification methods. Sinha and Sinha (2015) found that KNN outperforms SVM in accuracy, precision, and F measure, while SVM was better in calculating recall value. Vijayarani and Dhayanand (2015) found that BPN (ANN) outperformed SVM in terms of classification accuracy and execution time.

Chen et al. (2016) examined the feasibility of two fuzzy classifiers named FAOM and FuRES for CKD diagnosis and compared them to the linear classification method of Partial Least Squares Discriminant Analysis (PLS-DA). Sree and Ramesh (2020) examined several learning algorithms in regression analysis and classification modes, with SVM being selected as the best method for diagnosis. Ensemble classifiers achieved the highest accuracy for ESD. Some studies based their ML diagnosis methods delicately on background diseases, and the performance of more complicated models for KDD is investigated.

4.1.1. Underlying disease

Hypertension, diabetic nephropathy, and glomerulonephritis are the main causes of CKD. MLs have been used to diagnose CKD in patients with these diseases. Leung et al. (2013) examined the performance of six classification methods for detecting Diabetic Kidney Disease (DKD) in people with type 2 diabetes. They found that the SVM method had the highest accuracy for DKD diagnosis. Chimwayi et al. (2017) used classification and clustering methods to investigate the relationship between CKD and diabetes, finding good performance when biological and genetic input data is included in features space.

4.1.2. Ensemble methods

Combining machine learning and deep learning methods can improve predictive algorithms' performance. Studies have developed ensemble learning algorithms for various purposes, such as diagnosing kidney stones using a new technique combining classifiers. These algorithms have shown robustness, accuracy, and speed. For instance, Kazemi and Mirroshandel (2018) used a GA-based algorithm to combine classifiers, resulting in a robust model with over 97% accuracy.

Song et al. (2019) proposed a novel ensemble feature selection method for detecting kidney stone risk factors, comparing Gradient Boosting (GB), LASSO, and Deep Neural Networks (DNN) to achieve the highest accuracy of 95%. Bhaskar (2020) presented a computationally Correlational Neural Network learning model integrated with SVM for automated kidney disease diagnosis. The hybrid method showed superior execution time and accuracy.

4.2. End stage (acute) detection of CKD

Di Noia et al. (2013) developed a computer-aided software to predict IgA Nephropathy (IgAN) kidney disease health status. The hybrid EM-based ANNs can diagnose IgAN End Stage Kidney Disease (ESKD) in suspected patients online and offline. Experimental results show the method effectively assists experienced physicians in treating suspected patients.

4.3. CKD stage determination

Doctors use three methods to determine a patient's stage of kidney disease: estimating GFR using test data, using medical devices like ultrasound and MRI, and using online platforms like Google engine for eGFR data.

4.3.1. GFR-based estimation

The GFR test is the gold standard medical method for kidney disease (CKD), evaluating kidney function by determining fluid filtration per unit time. It helps nephrologists determine the severity of kidney disease and choose treatment plans. The eGFR level is classified into five stages: Stage 1, Stage 2, Stage 3, Stage 4, and Stage 5. The MDRD formula is used to calculate eGFR levels and report the severity of CKD, with low GFR indicating kidney damage and stage 1 indicating normal or increased GFR.

$$eGFR = 186 \times \text{Creatinine}^{-1.154} \times \text{Age}^{-0.202} \left(\frac{\text{ml/min}}{1.73\text{m}^2} \right)$$

Where, the Creatinine and Age are medical attributes. And for females, the acquired result is multiplied by a factor of 0.742.

Al-Hyari et al. (2013) developed a clinical decision support system using NB, ANNs, and DT for CRF diagnosis. The MLP method was used, and the system enables the diagnosis of CRF and determines the disease stage based on GFR. Kate et al. (2016) conducted a study comparing SVM, LR, DT, and NB classification algorithms for Acute Kidney Injury (AKI) disease. The results showed that all four methods worked well in detection and prediction, with the LR method showing superiority. The choice of an ML algorithm depends on whether the model is used for prevention or hospital management. Gopika and Vanitha (2017b) identified and clustered individuals with renal disease using KM, K-Median, and Fuzzy C-Means clustering methods. The proposed model determined the disease stage based on eGFR severity. The GFR test is used as a medical gold standard for KDD, and ML methods combined with the GFR test enhance the accuracy of KDD.

4.3.2. Biomedical devices and image processing

Medical devices are effective for direct diagnosis and acquiring initial data for kidney disease (CKD) diagnosis. Non-invasive image-based technologies like ultrasonography have gained popularity for CKD detection, reducing time, cost, and the likelihood of patients contracting other diseases. However, there are disadvantages such as speckle noises and difficulty in segmentation. Chen et al. (2014) developed a computer-aided system using computer vision and machine learning to diagnose CKD and its stages. They used the SVM method to diagnose CKD and its stage using eGFR-based experimental tests, achieving higher accuracy compared to eGFR tests. Jose et al. (2012) focused on segmenting ultrasonography kidney images for data processing and CKD stage identification using DM techniques. They extracted determining features to reduce DM complexity and classified three distinct kidney categories based on Bayesian theorem.

Table 2. Descriptive dimensions and coding.

Name	Description		Code
Diagnosis	Early-Stage Detection	The level or the severity of the disease for which the methodology is designed to diagnose and investigate.	ESD
	End Stage acute Detection		EnSD
Empty Cell	All Stage Detection		ASD
Methods	General roboroach for ML (e.g., classification, clustering, etc.)		Defined in Fig. 3 and appendix A.
Algorithms	Specific type(s) of used ML algorithms.		Defined in Fig. 3. and appendix A.
Data type	Real data	The UCI repository	RD-UCI
		Other sources	RD-OS
Empty Cell	Synthetic data		SD
CKD type distinction	e.g., KSD, Kidney ulceration, Kidney cysts, etc.	Yes	Y
		No	N
Validity evaluation	Validity measures like x-fold validation to test the authenticity and generalizability of classification methods.	Yes	Y
		No	N
Stage detection	Approach	GFR-based estimations	eGFR
		Biomedical device-based estimations	eBD
		Cloud-based platforms for self-detection	Online
Metrics	Accuracy evaluation measures like recall, accuracy, specificity, etc.		Defined in Table A in appendix B.

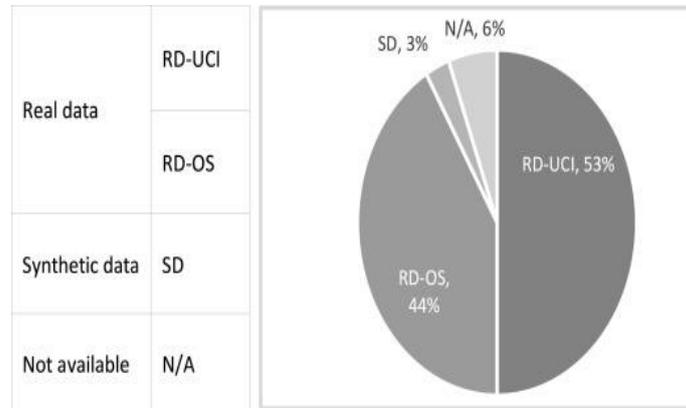


Fig.4. Data type percentage used in the literature

Table 3. Survey descriptive dimensions.

Reference	Year	Diagnosis	Methods	Algorithms	Metrics	CKD type distinction	Cross Validation	Stage detection	Data type
Eken et al.	2009	N/A	CIM	GA	AM, RSM, SSM	Y	Y	–	RD-OS
Abhishek et al.	2012	ESD	CM	BPN, RBF, SVM	AM, FM, PM, RSM, ETM, KSM, MAE, RAE	Y	N	–	RD-OS
Al-Hyari et al.	2013	ASD	CM	NB, DT, MLP	AM, RSM, SSM	N	Y	eGFR	RD-OS
Hsieh et al.	2014	ASD	Combined CMs	EM-based SVM	RSM, SSM	N	N	eBD	RD-OS
Chetty et al.	2015	ESD	DR (FeS)	WM, BFS	AM	N	N	–	RD-UCI
Empty Cell			CM	NB, SMO, KNN (IBK)					
Charleonnann et al.	2016	N/A	DR (FeS)	BFS	AM, SSM, RSM	N	Y	–	RD-UCI
Wibawa et al.	2017	ESD	DR (FeS)	CBFS	AM, PM, RSM, FM	N	Y	--	RD-UCI
Mohammed and Beshah	2018	ESD	CM	DT	AM	N	N	–	RD-OS
Saha et al.	2019	ESD	CM	SVM, RF, NB, MLP, LR	AM, FM, FM, RSM, RAE, KM, RMSE, RRSE, MAE	N	Y	–	RD-OS
Alaiad et al.	2020	ESD	AR	JRip, AA	AM, RSM, SSM, PM,	N	Y	–	RD-UCI

					MAE				
Al-Hyari et al.	2013	ASD	CM	NB, DT, MLP	AM, RSM, SSM	N	Y	eGFR	RD-OS
Hsieh et al.	2014	ASD	Combined CMs	EM-based SVM	RSM, SSM	N	N	eBD	RD-OS

4.3.3. Cloud-based applications

Internet-based applications like Google engine and MDRD formulation are used to determine the stage of kidney disease (CKD) using ANNs like BPN, MLP, and MNN. Chiu et al. (2012) found that MLP is superior for early detection of CKD risk. Ma et al. (2020) proposed a novel Deep Learning algorithm, Heterogeneous Modified ANN (HMANN), for early detection, segmentation, and diagnosis of CKD on the Internet of Medical Things (IoMT) platform. The HMANN algorithm achieved high accuracy in less execution time compared to traditional attribute selection methods.

4.4. Detecting different types of kidney diseases

Studies have limited knowledge on identifying specific types of kidney disease. Siwy et al. (2017) found that urinary proteome analysis can identify various types of kidney disease. Zheng et al. (2018) developed a hybrid classification algorithm using transfer learning and conventional imaging features for diagnosing kidney and urinary tract disease in children. Glazyrin et al. (2020) proposed a full-proteomic approach using plasma tests as a new biomarker for CKD type detection. The KNN algorithm was found to be the most effective in classifying CKD patients into groups with underlying diseases.

4.5. Descriptive dimensions

Table 2 contains the abbreviated codes of the various descriptive dimensions of the literature, which is mainly utilized in the descriptive tables of Section 4.5.1. and their related figures. The main purpose of this coding is to analyze various aspects of the literature with the most accuracy but in the shortest possible space.

4.5.1. Analysis of the literature

This section analyzes literature on kidney disease diagnosis using various methods, data types, and validation evaluation measures. Over 50% of studies use the UCI repository database, with 44% coming from other sources. A small percentage use synthetic data, especially for model testing after real data training. The literature review section provides detailed descriptions for each paper.

Table 4. Future research directions for KDD literature.

Name	Description
Utilizing DL tools for KDD	DL techniques are commonly used for image-oriented feature problems, but further studies are needed to fully explore their use for CKD prediction.
Utilizing DL tools for CKD stage detection	Utilizing deep learning techniques like convolutional neural networks can enhance the precision of identifying the stage of kidney disease.
Utilizing DL tools for CKD type determination	The use of deep learning algorithms can enhance the accuracy of diagnosing kidney disease.
Utilizing RL tools for KDD	RL methods have demonstrated satisfactory detection performance, but their application in KDD has been a persistent research gap.

Name	Description
Utilizing RL tools for CKD stage detection	The use of RL methods for disease stage determination through application, testing, and combination is expected to yield promising results.
Utilizing RL tools for CKD type determination	Few MLKDD literature studies have determined disease types, but no research has explored how RL methods work, which can yield significant results.
Online CKD detection using health applications on smartphones	The use of ensemble classifiers in the classification algorithm of health app data can potentially improve the diagnosis of CKD.
Ensemble classification-based methods with FeS for DR.	Ensemble classification-based methods with FeS for DR can enhance diagnostic learning algorithms' accuracy, providing effective tools for developing high-precision algorithms in KDD.
Data balancing	The SMOTE method, a novel approach in big data analysis, effectively reduces and analyzes healthcare data using the Hadoop framework and Map Reduce programming model.
GFR-based MLs	GFR test is a medical gold standard for MLs, but research gaps exist for testing GFR-based MLs and measuring evaluation metrics.

Fig. 5 shows that half of the studies on early detection of kidney disease (ESD) have been conducted, with the largest share being in determining ASD stage. The lowest number of studies is related to EnSD diagnosis, as the main goal is to determine if the disease is in the stage of kidney failure, which results in ESKD patients being placed on a dialysis list until a kidney transplant is provided.

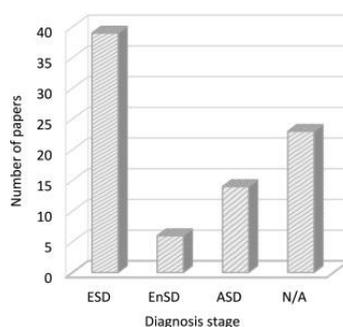


Fig. 5. The number of papers vs. diagnosis stage.

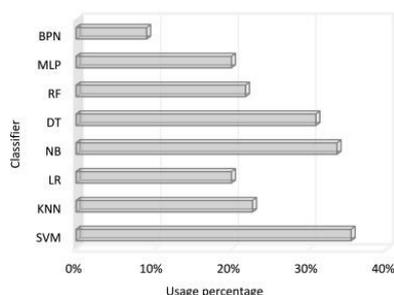


Fig. 6. The frequency of the most used classification algorithms

The most widely used algorithms in MLKDD are DT, SVM, and NB methods (Fig. 6), attributed to their greater kernel power and inherent feature of FeS. GFR estimation is the most common method for diagnosing kidney disease stage, as it is a medical gold standard and allows for classification in one of five stages. Biomedical devices like ultrasonography offer a lower-cost method for estimating disease stages. Two articles in this literature have developed internet-based software for remote sensing. The number of publications and stage diagnosis type also vary.

V. FUTURE RESEARCH DIRECTIONS

This section provides discovered research gaps for future works. We have provided most of our suggestions for future research in each section. Nonetheless, this section aims at summarizing related discussions and potential future research directions that are drawn from the overall review (see Table 4)

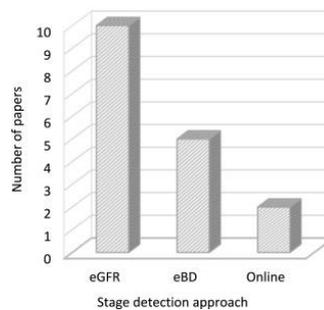


Fig. 7. CKD stage detection approach vs. frequency.

VI. CONCLUSIONS

This paper reviews the literature on the applications of machine learning for kidney disease diagnosis (MLD). It provides two comprehensive frameworks: one categorizes common learning algorithms used for KDD and the other proposes a framework for medical sub-fields related to MLD. The paper discusses the discoveries and novel contributions of machine learning (ML) for KDD, revealing that the efficiency and accuracy of ML methods for KDD depend on input data, data preparation, training dataset, and feature selection methods. Successfully implemented ensemble methods generally outperform routine MLs. The paper concludes that ML combined with the GFR test, a medical gold standard for KDD, enhances KDD accuracy compared to using either MLs or GFR tests alone. The paper also highlights research gaps and discusses future study directions, highlighting the lack of richer DL methods implementation for examining different aspects of kidney disease using computer vision for image segmentation.

REFERENCES

- [1] **Rajpurkar, P., et al. (2017).** CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. arXiv:1711.05225.
- [2] **Esteva, A., et al. (2019).** A guide to deep learning in healthcare. *Nature Medicine*, 25(1), 24–29.
- [3] **Chicco, D., & Jurman, G. (2020).** Machine learning in medicine: A review of algorithms, methods, and applications. *Journal of Medicine and Systems*, 44(1), 1-9.
- [4] **Soni, P., et al. (2019).** Machine Learning and Its Application in Kidney Disease Diagnosis: A Survey. *Computers in Biology and Medicine*, 115, 103495.
- [5] **Zhou, J., et al. (2020).** Machine learning techniques in the prediction and diagnosis of chronic kidney disease: A review. *Journal of Medical Systems*, 44(1), 1-9.

- [6] **Jiang, F., et al. (2017).** *Artificial intelligence in healthcare: Past, present and future.* Seminars in Cancer Biology, 53, 1-11.