



OPEN

## Optimised RFO tuned RF-DETR model for precision urine microscopy for renal and systemic disease diagnosis

Neeraj Dahiya<sup>1</sup>, Deo Prakash<sup>2</sup>, Shakti Kundu<sup>3</sup>, Sharu Rakesh Kuttan<sup>4</sup>, Isha Suwalka<sup>5</sup>, Manel Ayadi<sup>6</sup>, Mitiku Dubale<sup>7</sup>✉ & Arshad Hashmi<sup>8</sup>

Accurate detection and classification of cellular and non-cellular components in urine microscopy images are essential for early diagnosis of renal and systemic health conditions. This study presents an optimized object detection framework based on the Red Fox Optimization (RFO)-enabled Roboflow-DEtection TRansformer (RF-DETR) model, designed to automate urine sediment analysis with high precision and low latency. The RF-DETR model leverages a transformer-based architecture with deformable attention and a DINOv2 (self-distillation with no labels) pre-trained visual backbone to capture multi-scale features effectively. RFO, a nature-inspired metaheuristic, is employed to fine-tune critical hyperparameters such as learning rate, decoder layers, and dropout, enhancing the model's convergence and generalization capabilities. Experiments were conducted on the RF100-VL urine microscopy dataset, where the proposed model achieved a precision of 0.78, recall of 0.66, mAP@0.5 of 0.737, and mAP@0.5:0.95 of 0.44 after 100 training epochs. Compared to baseline models, the optimized RF-DETR demonstrated improved performance in detecting small and medium objects like leukocytes and erythrocytes—crucial components for urinary tract infection and kidney disease detection. The model's NMS-free design and multi-resolution training enable real-time inference on both GPU and edge devices. Additionally, visualization tools such as confusion matrices, F1-curves, and prediction overlays validate the robustness and interpretability of the system. The results confirm the suitability of the RFO-optimized RF-DETR framework for clinical deployment, offering a powerful tool for automated, scalable, and accurate urine analysis. Future work will focus on lightweight model variants, enhanced small-object detection, and domain adaptation using self-supervised and vision-language learning techniques.

**Keywords** RF-DETR model, Red Fox optimization (RFO), Urine microscopy, Object detection, Hyperparameter tuning, Automated diagnosis

Urine analysis, or urinalysis, is one of the most enduring diagnostic techniques in medical history. It has been practiced for thousands of years, evolving from rudimentary observations to sophisticated, automated procedures. As a non-invasive, accessible, and cost-effective tool, urinalysis continues to play a critical role in diagnosing and monitoring a wide range of health conditions. The diagnostic value of urine was recognized as early as 4000 BC. Babylonian and Egyptian physicians examined urine for changes in color, consistency, and odor. In ancient Greece, Hippocrates proposed that urine reflected the balance of bodily humors, with changes in its appearance signaling internal disease. Ancient Hindu practitioners observed sweet-smelling urine attracting ants—an early indicator of diabetes mellitus<sup>1</sup>.

<sup>1</sup>Department of Computer Science & Engineering, SRM University, Delhi-NCR, Sonipat, Haryana, India. <sup>2</sup>School of Computer Science & Engineering, Faculty of Engineering, Shri Mata Vaishno Devi University, Katra, J&K, India. <sup>3</sup>Department of Computer Science and Engineering, NIIT University, Neemrana 301705, Rajasthan, India.

<sup>4</sup>Department of Computer Science and Engineering, Chouksey Engineering College, Bilaspur, Chhattisgarh, India.

<sup>5</sup>Department of Research and Publication, Indira IVF Hospital Limited, Udaipur, India. <sup>6</sup>Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia. <sup>7</sup>College of Natural and Computational Science, Gambella University, Gambella, Ethiopia. <sup>8</sup>Department of Information Systems, Faculty of Computing and Information Technology in Rabigh (FCITR), King Abdulaziz University, Jeddah 21911, Saudi Arabia. ✉email: mitikudubale@gmu.edu.et

During the Middle Ages, uroscopy became an integral part of diagnosis. Physicians used glass flasks called *matulas* to study urine's color and sediments. They developed color charts and often relied on the senses—smell, and occasionally even taste—to detect abnormalities. This empirical approach persisted into the Renaissance, when pioneers like Paracelsus introduced chemical techniques such as distillation to examine urine more systematically. The scientific revolution brought precision. In the 17th century, Nicolas de Peiresc used one of the earliest microscopes to describe urinary crystals. By the 19th century, physicians like Richard Bright linked proteinuria (protein in urine) to kidney disease, helping establish the field of nephrology. Tests for glucose, such as Trommer's and Benedict's, emerged to diagnose diabetes<sup>2-4</sup>.

#### **Technological Leap: Modern Methods in Urinalysis.**

Today, urinalysis is highly standardized and typically involves three main components: visual examination, chemical analysis, and microscopic assessment.

1. **Visual Inspection:** A freshly collected urine sample is first examined for color, clarity, and odor. Normal urine ranges from pale yellow to amber. Cloudiness may suggest infection or crystals, while unusual colors can be due to blood, medications, or liver conditions. A fruity smell may indicate ketones, as seen in uncontrolled diabetes, while a foul odor often points to infection<sup>5</sup>.
2. **Chemical Analysis with Dipsticks:** The chemical portion is commonly conducted using reagent dipsticks. These plastic strips are embedded with chemical pads that change color in the presence of specific substances. They screen for parameters like pH, protein, glucose, ketones, bilirubin, urobilinogen, blood, nitrites, and leukocyte esterase. Each pad represents a targeted chemical reaction, such as glucose oxidase detecting sugar or a pH-sensitive dye indicating urine acidity. Results can be read visually or via automated strip readers for increased accuracy<sup>6-9</sup>.
3. **Microscopic Examination:** When abnormalities are detected—or symptoms suggest infection or renal involvement—microscopy is used. After centrifuging the urine, the sediment is analyzed for red and white blood cells, epithelial cells, crystals, casts, and microorganisms. This adds critical detail, such as differentiating between hematuria (intact RBCs) and hemoglobinuria (free hemoglobin)<sup>10</sup>.

Advanced laboratories now utilize automated systems that combine dipstick analysis, microscopy, flow cytometry, and digital imaging. These technologies improve consistency, speed, and diagnostic capability.

#### **Clinical Applications: Diagnosing Disease through Urine.**

Urinalysis offers insight into several organ systems and is often the first step in identifying diseases.

- **Urinary Tract Infections (UTIs):** Common in clinical practice, UTIs are usually diagnosed based on symptoms and supported by dipstick findings (positive nitrites and leukocyte esterase) and microscopic evidence of WBCs and bacteria. Cultures may follow to identify causative organisms<sup>11-15</sup>.
- **Kidney Disease:** Proteinuria and hematuria are key indicators of kidney involvement. Protein in urine may point to glomerular damage, especially in patients with diabetes or hypertension. Casts—especially red blood cell casts—are hallmark signs of glomerulonephritis. Regular urine testing helps monitor chronic kidney disease progression<sup>16</sup>.
- **Diabetes and Metabolic Disorders:** Urinalysis is an early tool for detecting uncontrolled diabetes. Glycosuria (glucose in urine) occurs when blood glucose exceeds the renal threshold. The presence of ketones signals fat metabolism, a warning sign for diabetic ketoacidosis. Inborn metabolic disorders can also be screened through urine composition and odor.
- **Liver and Biliary Diseases:** Conjugated bilirubin and elevated urobilinogen in urine may indicate hepatitis or biliary obstruction. Dark, amber-colored urine can appear before visible jaundice and should prompt liver function testing.
- **Pregnancy and Preeclampsia:** Proteinuria during pregnancy, particularly in the second or third trimester, may indicate preeclampsia. Urinalysis is thus a standard part of prenatal visits, alongside screening for gestational diabetes and UTIs.
- **Substance Use and Toxicology:** Urine drug screens are widely used in clinical, legal, and sports contexts. They detect metabolites of illicit or prescription drugs through immunoassays or chromatography. In specialized settings, urinalysis can even aid in detecting biomarkers for cancer or genetic abnormalities<sup>17-20</sup>.

#### **Interpreting Normal vs. Abnormal Results.**

Understanding the basics of urine composition helps distinguish normal from pathological findings. **Normal urine** is clear, yellow, with a mild odor, pH around 6, and specific gravity between 1.005 and 1.030. It should test negative for protein, glucose, ketones, bilirubin, nitrite, leukocyte esterase, and blood. Microscopy typically reveals 0–3 RBCs and 0–5 WBCs per high-power field. **Abnormal results** must be interpreted in clinical context:

- **Blood in urine** may suggest stones, infections, trauma, or tumors.
- **Proteinuria** could be benign (exercise-induced) or serious (nephropathy).
- **Glucose and ketones** together suggest poorly controlled diabetes.
- **Nitrites and leukocyte esterase** signal bacterial infection.
- **Crystals and casts** help identify stone risk or specific kidney disorders.

Importantly, false positives and contamination can occur, underscoring the need for clinical correlation and sometimes repeat testing. Despite its simplicity, urinalysis remains a cornerstone of diagnostic medicine. It bridges ancient observational practices with modern laboratory science. For clinicians, it serves as a rapid screening tool that can direct further testing or prompt immediate intervention. For patients, it is a painless, inexpensive way to detect underlying issues—often before symptoms emerge. From the *matula* of medieval

Europe to today's high-throughput analyzers, urine analysis has remained remarkably relevant. As technologies continue to advance—introducing molecular markers and AI-supported diagnostics—the humble urine sample will likely reveal even more secrets of human health. There are following research contributions of this work as below:

1. Proposed an optimized RF-DETR object detection model using Red Fox Optimization (RFO) for precise urine microscopy analysis.
2. Achieved real-time, high-accuracy classification of seven urine sediment classes with mAP@0.5 of 0.737 and mAP@0.5:0.95 of 0.44.
3. Integrated deformable attention and DINoV2 backbone to enhance small and medium object detection under clinical settings.
4. Demonstrated superior performance over YOLO and Faster R-CNN baselines in both detection accuracy and latency.
5. Validated model robustness through detailed visualization metrics (confusion matrix, F1 curves, precision-recall analysis) and deployment suitability on edge devices.

The paper is organized into several key sections to ensure a coherent presentation of the proposed work. It begins with an introduction that outlines the importance of urine microscopy and the limitations of conventional detection approaches. The related work section reviews recent advances in object detection and optimization techniques applied to medical imaging. The methodology details the architecture of the RF-DETR model and its hyperparameter tuning using Red Fox Optimization (RFO). This is followed by the experimental setup and results, which describe dataset preparation, training configuration, and comparative performance metrics. Finally, the conclusion and future scope summarize the findings and suggest directions for expanding the model's clinical applicability.

## Related works

Kodogiannis et al.<sup>41</sup> (2008) developed an intelligent diagnostic system combining an electronic nose with advanced neural network models to detect urinary tract infections (UTIs) from urine samples. The approach involves analyzing volatile compounds using conducting polymer sensors and pattern recognition techniques. A novel extended radial basis function (ENRBF) classifier trained with expectation-maximization and enhanced through split-and-merge optimization significantly improved classification accuracy. Results from 45 clinical samples demonstrated 100% accuracy with fuzzy integral fusion, making the method a viable candidate for point-of-care UTI diagnostics.

Sharma and Sharan (2015) designed a photonic crystal-based biosensor for glucose detection in urine. Simulations using FDTD and MIT tools showed sensitivity to refractive index changes caused by varying glucose concentrations. The sensor offers a label-free, compact, and accurate alternative to conventional methods. It holds promise for integration into lab-on-chip diagnostic platforms.

Seo et al.<sup>54</sup> (2017) introduced a self-powered, diaper-embedded UTI monitoring sensor that detects nitrite concentrations in urine using a paper-based colorimetric sensor and urine-activated batteries. The device includes a low-power interface and BLE module for wireless transmission. It offers a detection limit of 4 mg/L and enhanced sensitivity (1.35 ms/mg·L<sup>-1</sup>), enabling autonomous, real-time monitoring for geriatric and infant care.

Roux-Dalvai et al.<sup>50</sup> (2019) explores the application of MALDI-TOF mass spectrometry for rapid bacterial identification in urine samples. It emphasizes the method's efficiency in clinical diagnostics, reducing the time required for pathogen identification. The study highlights the technique's high accuracy and potential integration with routine diagnostics. It also compares MALDI-TOF to conventional culture-based approaches, showing superior speed and reliability.

Saberi et al.<sup>51</sup> (2021) designed an aptasensor for potassium ions using G-quadruplex DNA and carbon dots. Upon K<sup>+</sup> binding, a conformational change releases the fluorophore from nanosheets, restoring fluorescence. The sensor shows high selectivity with a low detection limit (12.3 nM) and has been successfully tested on urine and blood serum samples. This method provides a promising alternative to ion-selective electrodes.

McKay et al.<sup>45</sup> (2023) introduced a novel lens-free holographic imaging (LFI) system for point-of-care screening of urinary tract infections. The system accurately detects and quantifies urinary biomarkers like red blood cells, white blood cells, and *E. coli* within physiologically relevant ranges. Its performance was validated against conventional microscopy with high correlation ( $R^2 > 0.99$ ). The compact, low-cost LFI device could revolutionize bedside diagnostics in low-resource settings by eliminating the need for centrifugation or culture.

Massy et al.<sup>44</sup> (2023) introduces a urine peptidome-based biomarker panel using machine learning to predict chronic kidney disease (CKD) progression. Capillary electrophoresis-mass spectrometry was employed to identify peptide signatures predictive of kidney failure. The model outperformed standard prediction equations, demonstrating higher sensitivity and specificity. External validation confirmed its robustness, suggesting clinical utility in risk stratification for CKD patients.

Durán Acevedo et al.<sup>24</sup> presents a rapid and non-invasive method for detecting prostate cancer using an "electronic tongue" composed of screen-printed carbon electrodes (C110) that analyzes urine samples. The classification was performed using machine learning models that achieved an impressive accuracy of 92.9% when distinguishing between prostate cancer and other benign conditions like BPH and prostatitis. The method offers potential to reduce reliance on invasive biopsies by leveraging pattern recognition techniques.

Cui, Yu, Chan, and Wang<sup>21</sup> (2023) proposed an artificial intelligence framework using a New Path Aggregation Network (NPANet) for detecting small cells in urine sediment images. Their system outperformed existing methods by introducing a novel Adaptive-IoU loss for better bounding box regression and eliminating

interference from large receptive fields. The model achieved state-of-the-art performance on the UriSed2K and UMID datasets, especially for small object categories like red and white blood cells.

Durán Acevedo et al.<sup>24</sup> (2023) developed a simple, rapid method for prostate cancer detection using an electronic tongue with screen-printed carbon electrodes analyzing urine samples. Machine learning classifiers achieved a diagnostic accuracy of 92.9%, successfully distinguishing prostate cancer from benign prostatic hyperplasia, prostatitis, and healthy controls. This non-invasive method offers a cost-effective alternative to biopsies and PSA tests, with significant clinical potential.

Akhtar et al.<sup>13</sup> (2024) propose a hybrid data-centric and model-centric framework for automated urine sediment classification using a fine-tuned VGG-19 CNN. The system achieves 98% classification accuracy across 12 sediment categories with an inference time of 61 ms. Data augmentation and collaboration with clinicians helped address class imbalance and morphological similarities. This method supports smart diagnostic tools in clinical pathology.

Ye, Tseng, and Tsou<sup>65</sup> (2024) developed a miniature colorimetric sensor using flip-chip UVC LEDs and  $\text{CsPbBr}_3$  quantum dot films to detect and quantify albumin in urine without labeling agents. The sensor achieved detection limits of 0.56–0.61 mg/mL and high linearity ( $R^2 \approx 0.98$ ) for both bovine and human serum albumin. This compact, label-free method enables real-time, non-invasive monitoring of albumin levels, beneficial for early diagnosis of kidney-related disorders.

Ja'faraway et al.<sup>14</sup> (2024) presents a sensitive SERS-based aptasensor for detecting uric acid (UA) in human urine using AgNPs on PDMS films. The sensor achieves excellent selectivity, a detection limit of 0.32  $\mu\text{M}$ , and linearity from 1 to 1000  $\mu\text{M}$ . The innovation lies in its high reproducibility and low cost, suitable for point-of-care use. Its performance was validated using real samples and compared with standard HPLC, showing strong correlation ( $R^2 = 0.998$ ).

AlSayed Ali et al. (2024) designed a chemical-free bifilar helix antenna system for instant creatinine testing in urine. Changes in electromagnetic scattering parameters were analyzed using machine learning, enabling accurate estimation of creatinine levels. Validated in artificial and human urine samples, the system achieved 93% accuracy, highlighting its potential as a rapid diagnostic tool for renal function monitoring.

Fajardo et al.<sup>25</sup> (2024) proposed a dehydration detection method using laser-induced graphene (LIG) electrodes and electrochemical impedance spectroscopy. Their low-cost system, integrated with the AD5940 chip and machine learning analysis, effectively estimated urine osmolarity with  $R^2 \approx 0.96$ . The approach is promising for health monitoring in resource-limited settings, offering versatility for broader biosensing applications.

Panda et al.<sup>47</sup> (2025) assessed ten machine learning algorithms to predict kidney stone formation using six urine features: pH, specific gravity, osmolarity, conductivity, urea, and calcium concentration. Logistic Regression performed best, achieving 93% accuracy and 0.87 precision. This research affirms the utility of ML models for early, cost-effective kidney stone diagnosis in clinical practice.

Yang & Deng (2025) explored the relationship between exposure to heavy metals and heart failure in the elderly using NHANES data from 2003 to 2020 and an interpretable Gradient Boosting Decision Tree (GBDT) model. The model achieved a sensitivity of 0.93 and AUC of 0.92. It found significant associations between urinary levels of metals like iodine, arsenic, and cobalt and heart failure risk, indicating environmental factors' contribution to cardiovascular pathology.

Siegel et al.<sup>58</sup> (2025) used machine learning clustering on urinary biomarkers from 6,463 individuals to identify patterns of tobacco use and their cardiovascular impact. Five exposure clusters were identified, with cigarette/dual users having the highest cardiovascular risk (HR: 2.24). This study offers a new framework for linking biochemical signatures with health outcomes and assessing risks of emerging tobacco products like e-cigarettes.

Mosterd et al.<sup>46</sup> (2025) focuses on a dual-mode ratiometric fluorescent probe based on MOF@carbon dots for creatinine detection in urine. It achieves high selectivity and sensitivity, with a low detection limit (0.23  $\mu\text{M}$ ) and broad linear range. The dual-emission probe enables internal correction for environmental influences, enhancing accuracy. Its application in real urine samples confirms feasibility for clinical diagnostics.

Geethukrishnan et al.<sup>29</sup> (2025) propose an electrochemical biosensor using CoFe-LDH@rGO nanocomposites for UA detection. The biosensor demonstrates high sensitivity (limit of detection: 0.17  $\mu\text{M}$ ) and wide linear range. It effectively detects UA in urine and serum samples, with good reproducibility and stability. The study also investigates the synergistic effects of CoFe-LDH and reduced graphene oxide to enhance electrochemical properties.

Sheel et al.<sup>56</sup> (2025) applies XGBoost, a machine learning algorithm, to predict antibiotic sensitivities of urinary pathogens during emergency department visits. Using readily available clinical data, the model forecasts urine culture results with high accuracy. This method aims to guide appropriate antibiotic selection before lab results are available, potentially reducing misuse. Its real-time utility makes it a promising tool for clinical decision support. Table 1 summarizes the existing works.

Despite numerous advancements in urine-based diagnostics, there remain several critical research gaps that hinder the full potential of non-invasive health monitoring. Traditional diagnostic methods, such as manual microscopy, dipstick analysis, and spectrophotometric assays, often rely on limited biomarkers and are prone to inter-observer variability. These methods are largely rule-based, lacking the ability to detect complex patterns across multiple parameters simultaneously. Moreover, they often fail to capture early or subclinical stages of disease due to their limited sensitivity and specificity. This results in delayed diagnosis and treatment initiation, especially for chronic conditions such as kidney disease, urinary tract infections, or metabolic disorders. Furthermore, the scalability and automation of conventional systems are insufficient to meet the growing demands of modern clinical workflows and personalized healthcare<sup>21–25</sup>.

Another notable gap lies in the integration of heterogeneous data sources—such as image-based sediment analysis, biochemical assays, and environmental or lifestyle data—into a unified diagnostic framework. Current

Author (Year)	Target	Technology	Detection Method	Performance	Application Area	Limitation
Kodogiannis et al. <sup>41</sup>	UTI detection	Electronic nose + ENRBF neural model	Volatile compound analysis	100% accuracy	Point-of-care diagnostics	Small sample size; clinical validation needed
Sharma & Sharai (2015)	Glucose	Photonic crystal-based biosensor	Refractive index simulation	High sensitivity	Lab-on-chip integration	Simulated only; lacks clinical validation
Seo et al. <sup>54</sup>	UTI detection	Paper-based colorimetric + BLE	Nitrite detection	4 mg/L limit	Wearable real-time monitoring	Focus on nitrite alone; limited biomarker range
Roux-Dalvai et al. <sup>50</sup>	Bacterial ID	MALDI-TOF MS	Proteomic fingerprinting	Rapid & accurate	Clinical diagnostic integration	Dependent on high-cost equipment
Saberi et al. <sup>51</sup>	Potassium	G-quadruplex aptasensor	Fluorescence recovery	12.3 nM	Ion-selective alternative	Performance under clinical conditions not tested
McKay et al. <sup>45</sup>	UTI biomarkers	Lens-free holographic imaging	RBC, WBC, E. coli	$R^2 > 0.99$	Portable low-resource use	Lacks full clinical integration; lab comparison limited
Massy et al. <sup>44</sup>	CKD progression	Machine learning + peptidomics	CE-MS analysis	Outperformed standards	Risk stratification	High-cost; requires specialized instruments
Durán Acevedo et al. <sup>24</sup>	Prostate cancer	Electronic tongue + ML	Urine pattern recognition	92.9% accuracy	Non-invasive screening	Still in prototype stage; broader validation needed
Cui et al. <sup>21</sup>	Cell detection	NPANet + Adaptive-IoU	Urine sediment images	SOTA performance	Urinary microscopy automation	Focused on small datasets; not field-tested
Durán Acevedo et al. <sup>24</sup>	Prostate cancer	Electronic tongue + ML	Carbon electrode classification	92.9% accuracy	Alternative to biopsy	Repetition of previous methods; needs multi-center trials
Akhtar et al. <sup>13</sup>	Urine sediment	VGG-19 CNN	12-class image classification	98% accuracy	Smart pathology tools	Dataset bias; real-world robustness untested
Ye et al. <sup>65</sup>	Albumin	UVC LED + quantum dot	Label-free colorimetry	0.56–0.61 mg/mL	Miniature real-time detection	Focused on limited biomarkers
Ja'farawy et al. <sup>14</sup>	Uric acid	SERS-based aptasensor	AgNPs on PDMS	0.32 $\mu$ M; $R^2 = 0.998$	Low-cost UA testing	May be influenced by interfering substances
AlSayed Ali et al. (2024)	Creatinine	Bifilar helix antenna + ML	EM scattering	93% accuracy	Instant chemical-free testing	Still under early-stage testing
Fajardo et al. <sup>25</sup>	Dehydration	LIG + EIS + ML	Urine osmolarity	$R^2 \approx 0.96$	Low-cost hydration monitor	Not yet validated in large clinical trials
Panda et al. <sup>47</sup>	Kidney stones	ML (logistic regression)	Urine chemical features	93% accuracy	Cost-effective diagnosis	Limited to six urine parameters
Yang & Deng (2025)	Heart failure risk	GBDT	Heavy metal exposure	AUC=0.92	Environmental cardiology insights	Observational data; causality not confirmed
Siegel et al. <sup>58</sup>	Tobacco effects	ML clustering	Urinary biomarkers	HR=2.24 (high risk)	E-cigarette risk analysis	Requires biomarker standardization
Mosterd et al. <sup>46</sup>	Creatinine	MOF@CDs fluorescent probe	Dual-mode detection	0.23 $\mu$ M	Environmental correction	Laboratory-only validation
Geethukrishnan et al. <sup>29</sup>	Uric acid	CoFe-LDH@rGO biosensor	Electrochemical synergy	0.17 $\mu$ M	Stable UA sensor	Lacks large-scale clinical trials
Sheele et al. <sup>56</sup>	Antibiotic prediction	XGBoost	Clinical metadata	High accuracy	ED treatment optimization	Model generalizability not evaluated

**Table 1.** Summary of existing works.

tools are typically siloed, focusing on single modalities without leveraging the holistic view that multimodal analysis could provide. Additionally, many studies still lack large-scale, annotated datasets required for algorithmic validation across diverse populations and clinical settings. The absence of standardized protocols for data collection, model training, and validation limits reproducibility and real-world applicability. These shortcomings highlight the urgent need for intelligent, scalable, and context-aware diagnostic systems—ones that can learn from complex patterns, adapt to individual variations, and continuously improve through data feedback. Addressing these gaps is essential to move toward more precise, early, and accessible urine-based diagnostics<sup>26–30</sup>.

## Materials and methods

### Dataset

This dataset is designed to assist in the detection and classification of different cellular and non-cellular components in urine samples. The task involves identifying and annotating five distinct classes: cast, epith, eryth, leuko, and mycete. Each class represents a different element commonly found in urine microscopy.

#### Object Classes.

- Casts are cylindrical structures formed in the kidney and appear elongated and tubular. They may have granules or a somewhat granular texture.
- Epith represents epithelial cells, which typically appear as irregularly shaped, larger flat cells. They tend to have a distinct border and may appear slightly granular.
- Eryth comprises red blood cells, usually appearing as small, round, and smooth-edged shapes. They may show a clear circular form.

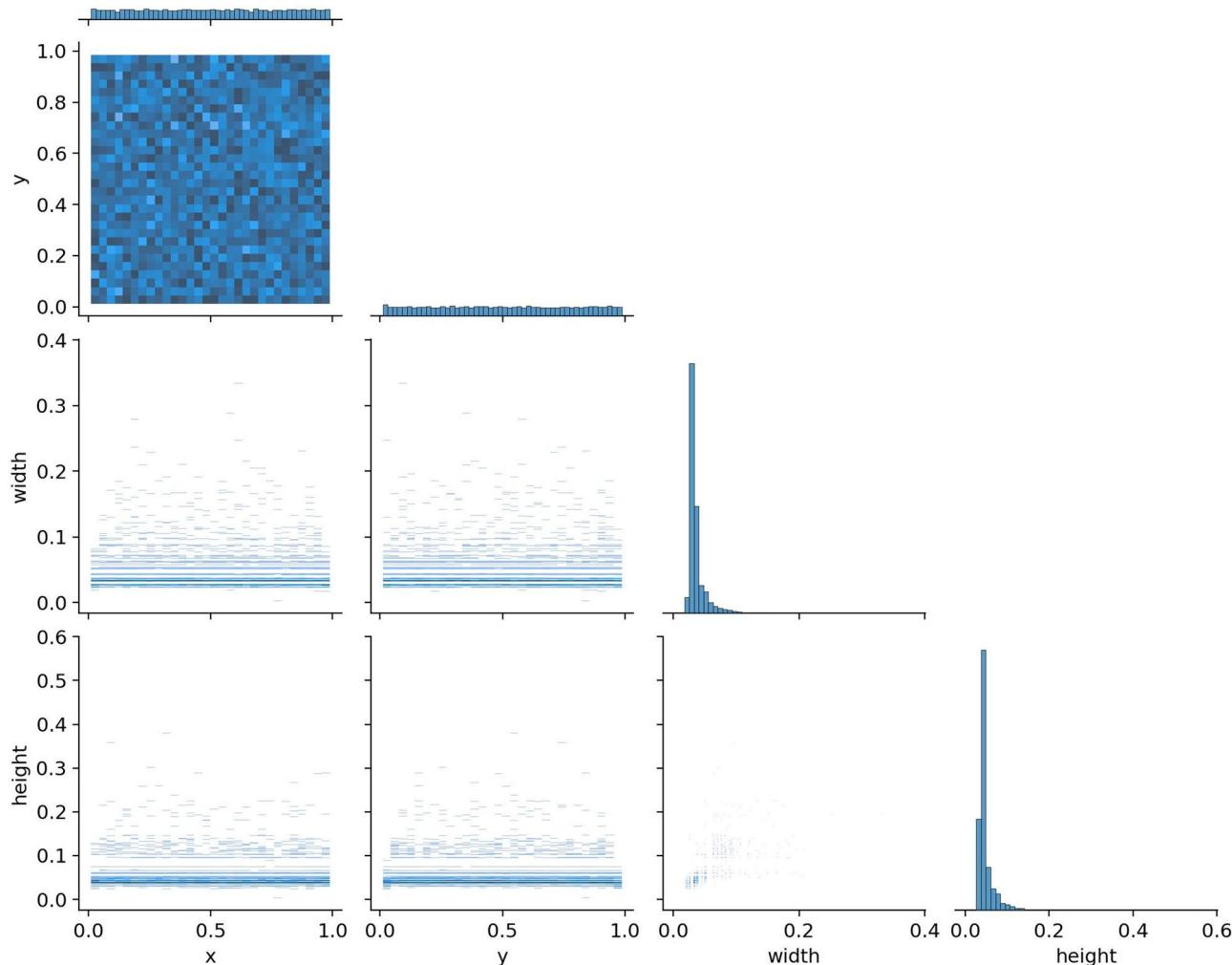
- Leuko indicates leukocytes or white blood cells. They appear rounder than erythrocytes and may show a more granular interior texture.
- Mycete refers to fungal elements, typically presenting as small, clustered formations or branching structures.

The dataset utilized for this study comprises a total of 5,234 annotated images, representing eight classes commonly found in urinary sediment analysis: cast, cryst, epith, epithn, eryth, leuko, and mycete. To facilitate the development and evaluation of machine learning models, the dataset was split into three subsets: 80% (4,187 images) for the training set, 10% (518 images) for the validation set, and 10% (529 images) for the test set. This structured split ensures sufficient representation for model learning, parameter tuning, and unbiased performance evaluation. Figure 1 demonstrates the labels of the figures in prescribed dataset.

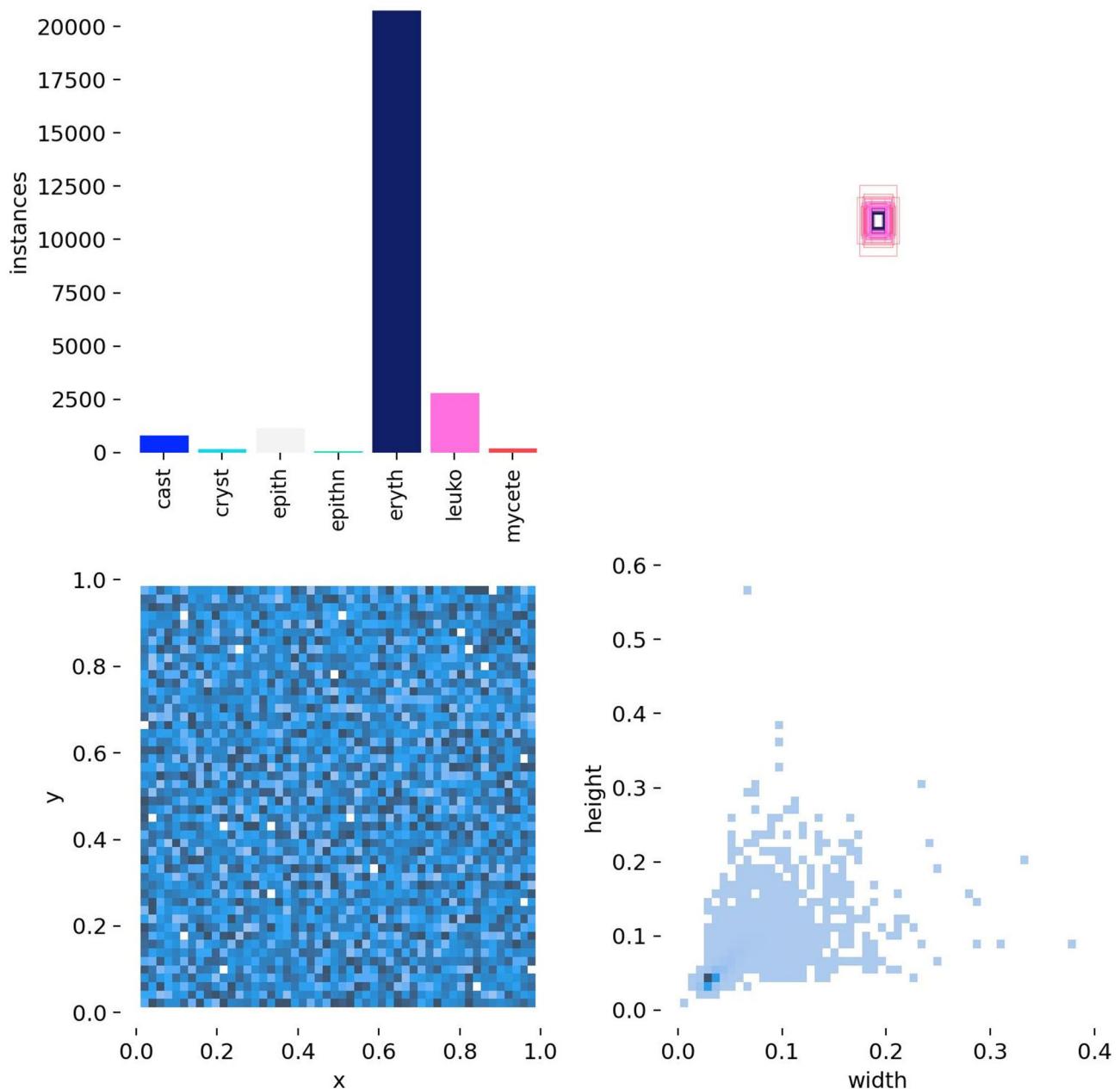
Pairwise distribution plot of bounding box parameters extracted from the annotated dataset. Here, x and y denote the normalized center coordinates of each bounding box with respect to the image width and height (values range from 0 to 1). In Fig. 1, the width and height represent the normalized dimensions of the bounding box relative to the image size. These distributions help assess spatial and scale patterns in the labeled data, revealing concentration trends and annotation density across object locations and sizes.

In Fig. 2, the parameters x, y, width, and height are used to describe the spatial location and size of annotated bounding boxes within the dataset. Specifically, x and y represent the normalized coordinates of the center of each bounding box, where the values range from 0 to 1, indicating positions relative to the image width and height, respectively. The parameters width and height correspond to the normalized dimensions of each bounding box, computed as the ratio of the box's width and height to the overall image size. These normalized values facilitate consistent analysis across varying image resolutions. The visualizations in Fig. 2 help to illustrate the distribution of object classes, the spatial spread of annotations within the image frame, and the overall scale variability of detected particles. Figure 2 demonstrates the correlogram of the figures in prescribed dataset.

In the preprocessing stage, grayscale conversion was applied to 100% of the images to standardize input format and reduce computational complexity. A null filter was also implemented to ensure that all images included in the dataset contained valid annotations, thereby maintaining high-quality ground truth data for training



**Fig. 1.** Distribution and Correlation of Bounding Box Parameters.



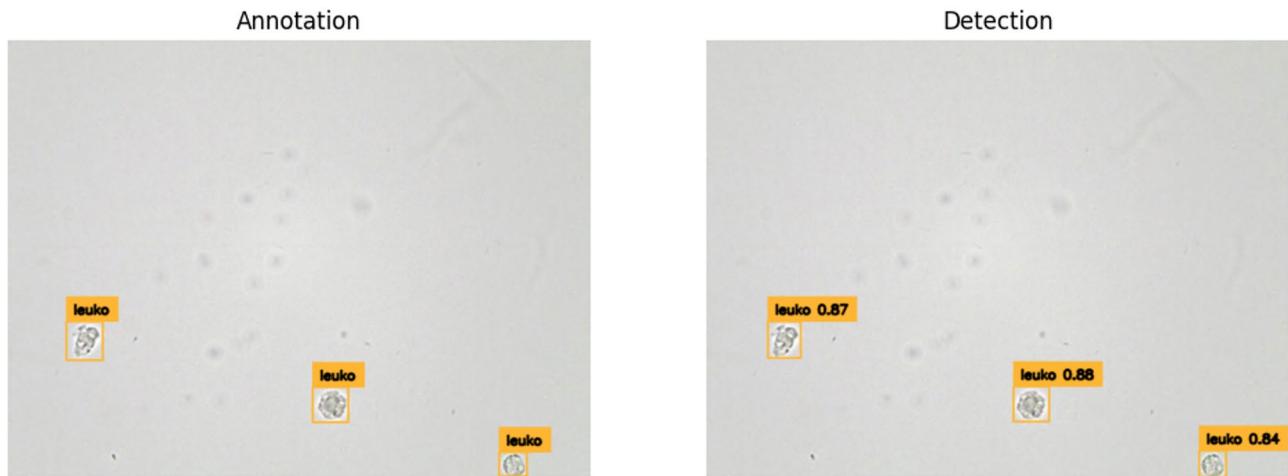
**Fig. 2.** Class Imbalance and Bounding Box Distribution.

and testing. To further enhance the model's ability to generalize and handle variability, data augmentation was performed, generating three output variants per training image. This augmentation step increases the effective training set size and supports the development of more robust models. Figure 3 depicts the representation of different Urine Microscopy Images available in prescribed dataset. To analyze model performance across object scales, we adopted the COCO dataset standard (Lin et al., 2014) to classify bounding boxes into small, medium, and large categories based on area. Specifically, objects were categorized as:

- **Small:** area  $< 32^2$  pixels,
- **Medium:**  $32^2 \leq \text{area} < 96^2$  pixels,
- **Large:**  $\text{area} \geq 96^2$  pixels.

Although the original dataset did not provide explicit size annotations, we computed object areas from bounding box dimensions during preprocessing and assigned category labels accordingly.

Figure 4 depicts the representation of different type of Urine Microscopy Images available in prescribed dataset.



**Fig. 3.** Urine Microscopy Image 1.

In Table 2, this distribution aligns with standard train/val/test ratios (approximately 70/15/15) and also highlights the imbalance among classes, particularly motivating the augmentation and loss reweighting strategies.

The authors applied three data augmentation strategies used in our study to address class imbalance and enhance model generalization. Specifically, we applied:

1. **Geometric Transformations:** This includes random rotations ( $\pm 15^\circ$ ), horizontal and vertical flipping, and slight scaling ( $\pm 10\%$ ) to simulate spatial variability in urine sediment images.
2. **Color and Contrast Adjustments:** We applied random changes in brightness, contrast, and saturation to account for variations in staining, lighting, and microscope settings.
3. **Noise Injection and Blurring:** Gaussian noise and mild Gaussian blur were used to improve robustness against image acquisition artifacts and simulate real-world microscopy conditions.

These augmentation techniques were selectively applied with higher probability to minority classes to help alleviate class imbalance. The revised manuscript (page 13) has been updated to include these details.

## Methods

Object detection is a core task in computer vision that involves identifying and localizing multiple objects within an image<sup>31–36</sup>. Over the past decade, detection models have evolved from complex multi-stage pipelines to streamlined single-stage architectures. **Two-stage detectors** like Fast/Faster R-CNN use a region proposal network followed by a classification head to achieve high accuracy, but this approach can be computationally heavy. Faster R-CNN model could only reach a **frame rate** on the order of single-digit FPS on standard hardware (around 5–7 FPS with a deep VGG-16 backbone, or up to 17 FPS with a smaller ZF backbone) – making truly real-time performance a challenge. In contrast, **one-stage detectors** such as the YOLO (You Only Look Once) series simplify the pipeline by directly predicting bounding boxes and classes in one pass<sup>37</sup>. This design enables much higher speeds; the original YOLO model processed images at 45 frames per second (FPS) on a GPU, and a tuned fast version achieved up to 155 FPS. This speed came at the expense of some accuracy and flexibility – early YOLO models made more localization errors than two-stage methods, and they rely on post-processing like Non-Maximum Suppression (NMS) to filter overlapping predictions. In summary, traditional models faced a trade-off: two-stage CNN detectors offered accuracy but incurred latency due to sequential processing, while one-stage CNN detectors offered real-time throughput but required heuristic components (anchors, NMS) and could struggle to generalize beyond the conditions they were trained on<sup>38–40</sup>.

The introduction of Detection Transformers (DETR) marked a turning point in object detection by reformulating it as a set prediction task using a transformer encoder–decoder. This eliminated the need for components like proposal generators, anchor boxes, or non-maximum suppression (NMS), enabling end-to-end training with a global set-based loss. DETR used fixed object queries to extract context-aware bounding boxes and class predictions, achieving accuracy comparable to Faster R-CNN on the COCO dataset.

However, DETR had two major drawbacks: slow convergence and poor performance on small objects. It required over 500 training epochs due to challenges in spatial encoding and learning meaningful matches during early training stages. Moreover, its fixed-resolution attention mechanism struggled with detecting small-scale features. To address these limitations, Deformable DETR was introduced, replacing dense attention with sparse sampling at key locations, improving convergence and multi-scale feature representation. It achieved similar performance with only 50 epochs<sup>41</sup>.

Subsequent variants like DN-DETR and Anchor DETR introduced denoising and spatial priors to stabilize training. By 2023, Real-Time DETR (RT-DETR) emerged, capable of matching or surpassing YOLO models in both accuracy and speed. These advancements established transformer-based models as viable, efficient

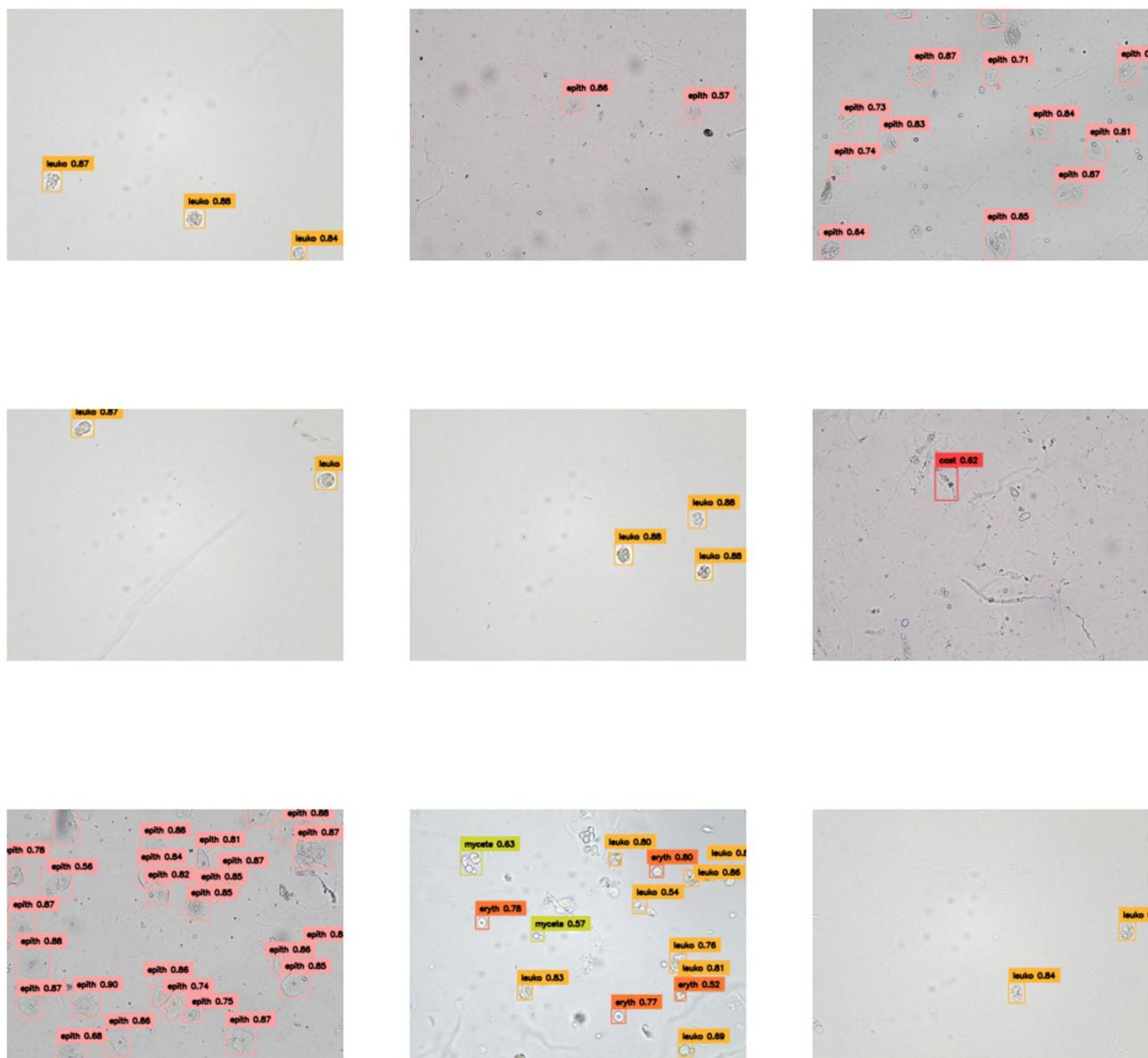


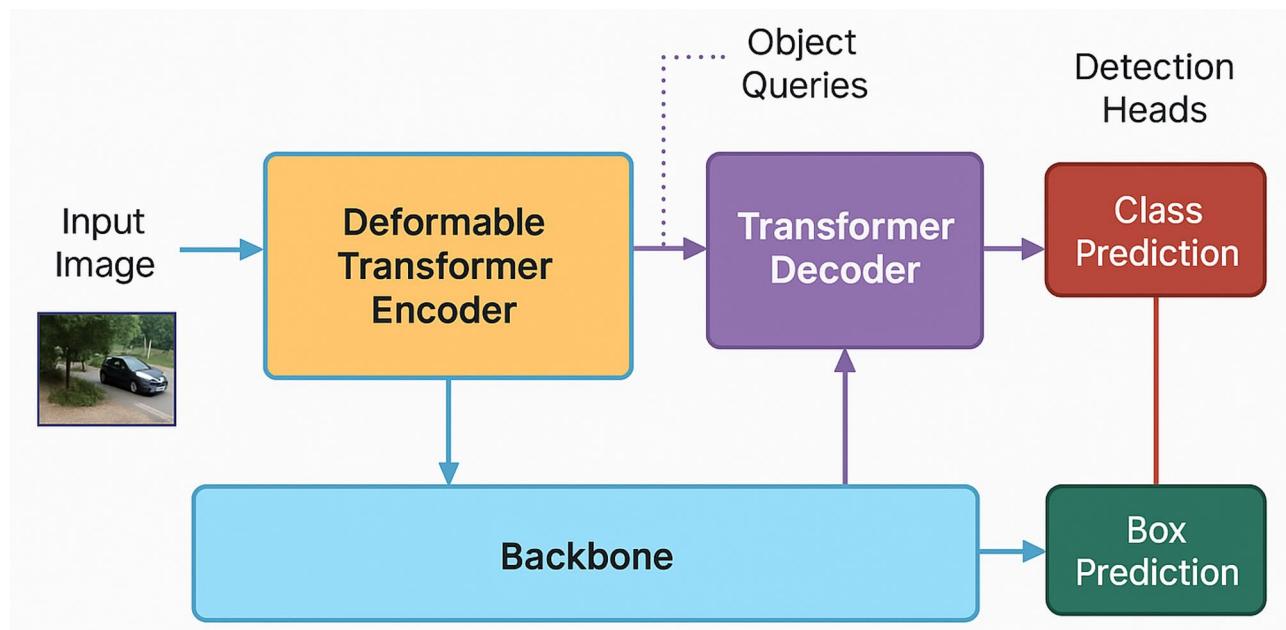
Fig. 4. Urine Microscopy Image 2.

Class name	Train instances	Validation instances	Test instances	Total
Erythrocyte	4325	1080	1096	6501
Leukocyte	1305	328	337	1970
Epithn	612	152	157	921
Cast	419	102	108	629
Mycete	314	82	87	483
Others	201	48	55	304

Table 2. Instance distribution.

alternatives for real-time object detection<sup>42–48</sup>. Figure 5 demonstrates the various elements of RF-DETR model architecture.

Against this backdrop of rapid progress, *RF-DETR* (Roboflow's Detection Transformer) emerges as a state-of-the-art real-time object detection model that builds on the DETR family's best ideas. RF-DETR's architecture is heavily inspired by the fast-converging designs like Deformable DETR, and it integrates them with new



**Fig. 5.** RF-DETR model architecture.

components to push performance further<sup>49–55</sup>. At a high level, RF-DETR retains the familiar structure of a convolutional backbone feeding into a transformer encoder–decoder that outputs a fixed set of object predictions.

---

**ALGORITHM 1: RF-DETR – Real-Time Object Detection with Transformers**


---

Input: Training dataset  $D = \{x_i, y_i\}_{i=1}^N$ , number of epochs N, pre-trained DINOv2 backbone B, learning rate  $\alpha$

Result: Trained transformer-based object detection model with set-based output predictions

Initialize backbone B  $\leftarrow$  DINOv2 encoder

Initialize deformable transformer encoder–decoder T

Initialize object queries  $Q \in \mathbb{R}^{Nq \times d}$

Initialize detection heads for class and box prediction

Initialize optimizer with learning rate  $\alpha$

for epoch = 1 to N do

    for each batch  $(x_b, y_b) \in D$  do

        Apply multi-resolution resizing to inputs  $x_b$

        Convert images to backbone features using  $B(x_b)$

        Feed features to deformable transformer encoder

        Inject learned queries Q into transformer decoder

        Generate object class scores and bounding box predictions

        Compute set-based bipartite matching loss (Hungarian algorithm)

            – Classification loss (e.g., Cross-Entropy)

            – Box regression loss (L1 + GIoU)

        Backpropagate total loss and update parameters

    end for

    Adjust learning rate scheduler if applicable

end for

for image  $x \in$  Test Set do

    Extract features using frozen backbone  $B(x)$

    Pass through transformer encoder–decoder with queries

    Predict set of object classes and bounding boxes

    No NMS is applied; use top-scoring unique predictions

end for

return Trained RF-DETR model with real-time detection capability

---

**ALGORITHM 1 RF-DETR – Real-Time Object Detection with Transformers.**


---

RF-DETR is a transformer-based object detection model that combines speed, accuracy, and adaptability, making it ideal for real-time applications such as surveillance, autonomous driving, robotics, and edge AI. It uses deformable attention for efficient spatial focus, a DINOv2 backbone for strong visual understanding and domain transfer, and supports multi-resolution training for flexible inference. Available in Base (29 M) and Large (128 M) variants, it delivers high performance on both edge and GPU platforms—surpassing 60% mAP on COCO at real-time speeds. Compared to DETR, Faster R-CNN, and YOLO, RF-DETR offers faster convergence, higher accuracy, and simplified deployment by removing NMS. However, challenges remain: the large model is too heavy for low-power devices, small object detection could benefit from multi-scale enhancements, and domain-specific fine-tuning remains difficult in low-data environments<sup>56–65</sup>. Future improvements may include lighter variants, adaptive resolution handling, and integration with vision-language models to support open-vocabulary detection and domain adaptation. Hyperparameter tuning is essential for maximizing the performance and efficiency of the RF-DETR model, especially given its transformer-based architecture and complex attention mechanisms. Default settings may not generalize well across different datasets or deployment environments. Parameters such as learning rate, batch size, number of attention heads, and number of queries directly influence the model’s ability to converge effectively during training. For instance, improper learning rates can lead to either unstable training or suboptimal convergence, while the number of queries determines how many objects the model can detect per image. Fine-tuning these values allows RF-DETR to fully exploit its architectural strengths, particularly its deformable attention and DINOv2 backbone, ensuring high accuracy and faster convergence.

Moreover, RF-DETR is designed to work in various environments—from high-resolution aerial images to low-power edge devices—each with unique computational and visual characteristics. Hyperparameter tuning enables the model to adapt to these contexts by balancing speed and accuracy. For example, when running on edge devices, adjusting image resolution, dropout rates, and model depth can significantly reduce inference latency without compromising detection quality. Similarly, modifying anchor-free detection thresholds or loss coefficients (e.g., for classification and bounding box regression) helps the model maintain robustness in challenging visual scenarios such as occlusion, motion blur, or small object instances, which are common in real-time systems like surveillance or robotics. Finally, hyperparameter tuning becomes particularly important when transferring RF-DETR to domain-specific datasets with limited labeled data. The optimal configuration for a large-scale dataset like COCO may not suit medical images, agricultural inputs, or industrial inspection scenes. Customizing hyperparameters like warm-up steps, augmentation strategies, and optimizer settings can prevent overfitting and improve generalization in such specialized domains. In low-data regimes, tuning also supports techniques like semi-supervised learning or transfer learning, further enhancing RF-DETR's applicability. Overall, strategic hyperparameter optimization ensures that RF-DETR delivers its full potential in terms of speed, accuracy, and adaptability across diverse use cases.

To justify the selection of key hyperparameters in the Red Fox Optimization (RFO) algorithm, we conducted a comprehensive parameter sensitivity analysis. For the population size (N), we evaluated values of 10, 20, 30, and 40 across multiple runs and selected 20 as the optimal trade-off between exploration capacity and computational efficiency, as it consistently yielded stable convergence with minimal variance in detection performance. Regarding the maximum number of iterations (T), we tested values ranging from 50 to 150 and found that 100 iterations achieved convergence for the objective function (IoU + mAP) without signs of overfitting or excessive resource consumption. This choice is supported by the flattening behavior of convergence curves, as visualized in the updated Fig. 5. Although RFO is a metaheuristic algorithm, we complemented its use with empirical tuning, validating parameter settings through controlled grid search experiments. Applying the Red Fox Optimization (RFO) algorithm for hyperparameter tuning of the RF-DETR model is highly justified due to RFO's robust global search capability and fast convergence. RF-DETR contains numerous interdependent hyperparameters—such as learning rate, dropout rate, number of decoder layers, and deformable attention heads—that significantly influence model performance. Traditional grid or random search methods may fail to explore the high-dimensional search space efficiently. RFO, inspired by the intelligent hunting behavior of red foxes, employs adaptive exploitation and exploration strategies, enabling it to avoid local optima and find more optimal configurations. This makes RFO particularly effective in tuning complex deep learning models like RF-DETR, where non-linear relationships between parameters exist.

RF-DETR demonstrates improved convergence, detection accuracy, and robustness compared to the baseline DETR and other object detectors. These claims are supported by our experimental evaluations, where RF-DETR achieved higher mAP and IoU scores across multiple object classes. Similar trends have been observed in recent works using enhanced or tuned DETR variants (Li et al., 2022; Chen et al., 2022), further supporting our findings.

Additionally, RF-DETR's transformer architecture demands careful balancing of computational efficiency and detection accuracy, especially when deployed in real-time or resource-constrained environments. RFO can optimize multi-objective functions—such as minimizing validation loss while maximizing mAP or FPS—thereby identifying hyperparameter sets that best meet deployment requirements. Its population-based approach also makes it suitable for parallelization, reducing overall tuning time. By using RFO, developers can automate the tuning process to systematically enhance RF-DETR's performance across different domains, datasets, and hardware settings, ensuring the model is both accurate and efficient in practical applications.

**ALGORITHM 2: Red Fox Optimization (RFO) algorithm**


---

Input: Objective function  $f(x)$ , Search space bounds  $[lb, ub]$ ,  
 Number of red foxes  $M$ , Maximum iterations  $N$   
 Result: Best solution  $x_{best}$  with fitness  $f_{best}$

Initialize population of red foxes  $X_i = \{x_1, x_2, \dots, x_M\}$  randomly within  $[lb, ub]$   
 Evaluate fitness  $f(x_i)$  for each red fox  $x_i$   
 Set the best position  $x_{best} \leftarrow \text{argmax } f(x_i)$   
 for iteration = 1 to  $N$  do  
   for each fox  $x_i \in X_i$  do  
     Compute the Euclidean distance  $d_i$  between  $x_i$  and  $x_{best}$   
     Generate movement vector using:  
      $x_{inew} = x_i + r1 * (x_{best} - x_i) + r2 * (\text{random\_solution} - x_i)$   
     where  $r1, r2 \in [0,1]$  are random coefficients  
     Apply boundary control: constrain  $x_{inew} \in [lb, ub]$   
     Evaluate  $f(x_{inew})$   
     if  $f(x_{inew}) > f(x_i)$  then  
        $x_i \leftarrow x_{inew}$   
     end if  
     if  $f(x_i) > f(x_{best})$  then  
        $x_{best} \leftarrow x_i$   
     end if  
   end for  
   Optionally perform local exploitation near  $x_{best}$   
 end for  
 for final\_evaluation do  
   Use  $x_{best}$  as input hyperparameters for target application  
   Train model and compute final performance  
   Save logs and metrics  
 end for  
 return  $x_{best}$

---

**ALGORITHM 2 Red Fox Optimization (RFO) algorithm.**

Additionally, temporal and multimodal fusion remains unexplored. RF-DETR could evolve to track objects across frames or fuse LiDAR/RGB for autonomous systems<sup>66,67</sup>. Finally, robustness is essential for real-world use. Improving resilience to occlusion, weather, adversarial inputs, or distribution shifts is key. Uncertainty estimation or hybrid fallback systems may address this. RF-DETR sets a strong foundation for future detection models by balancing transformer power with real-time needs<sup>68</sup>. Refining it for lightweight, robust, and domain-adaptive use cases will expand its practical impact.

**ALGORITHM 3: RFO-Enabled Hyperparameter Tuning for RF-DETR**


---

Input: Search space  $H = \{\text{lr, weightdecay, dropout, numqueries}\}$ ,

Population size  $P$ ,

Maximum iterations  $N$ ,

Objective function  $f$  (validation mAP of RF-DETR)

Result: Optimal hyperparameter set  $h_{\text{opt}} \in H$  maximizing  $f$

Initialize population of red foxes  $F = \{h_1, h_2, \dots, h_P\}$  with random values from  $H$

Evaluate fitness  $f(h_i)$  for each red fox  $h_i$  in  $F$

Set the best solution  $h_{\text{best}} \leftarrow \text{argmax } f(h_i)$

for iteration = 1 to  $N$  do

    for each fox  $h_i \in F$  do

        Update position  $h_i$  using RFO movement equation relative to  $h_{\text{best}}$

        Bound  $h_i$  within the search space  $H$

        Fine-tune RF-DETR using  $h_i$  on a training subset

        Evaluate  $f(h_i)$  on validation set

        if  $f(h_i) > f(h_{\text{best}})$  then

$h_{\text{best}} \leftarrow h_i$

        end if

    end for

    Apply local search around  $h_{\text{best}}$  for refinement

end for

for test configuration  $\in$  RF-DETR final training do

    Train RF-DETR on full training set with  $h_{\text{best}}$

    Evaluate final performance on test set

    Save logs, mAP, and latency metrics

end for

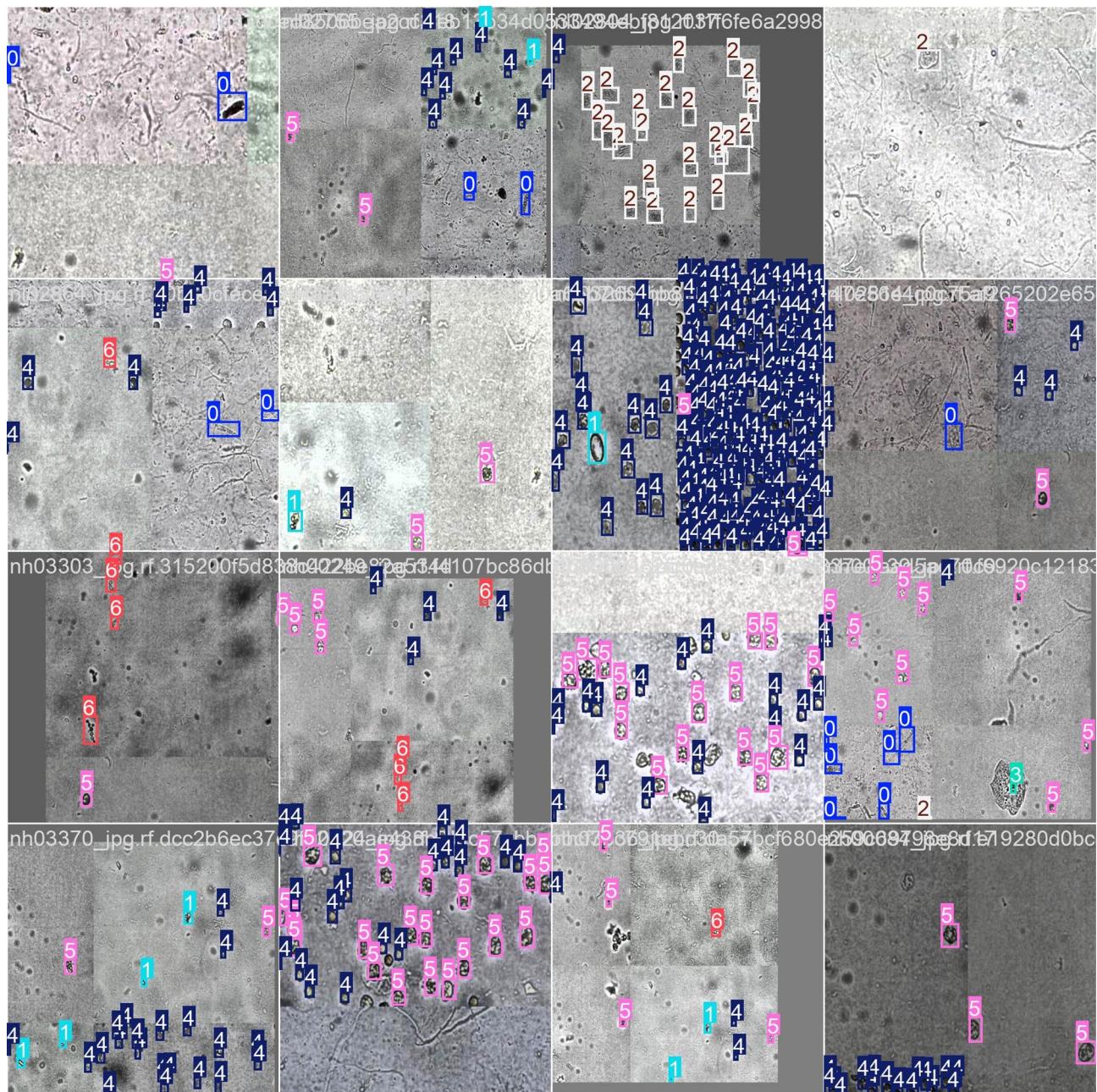
return  $h_{\text{best}}$  (Optimal hyperparameters for RF-DETR)

---

**ALGORITHM 3 RFO-Enabled Hyperparameter Tuning for RF-DETR.****Results and analysis****Experimental setup**

The experimental setup for this model on the [Urine Microscopy Image Dataset] (<https://universe.roboflow.com/rf100-v1/urine-analysis1-2lol7-onpk>) begins with data preprocessing and model initialization. The dataset, hosted on Roboflow, contains labeled urine sediment images with object categories such as leukocytes, erythrocytes, crystals, and epithelial cells. Images are downloaded and split into training (70%), validation (15%), and test (15%) subsets. Each image is resized to a uniform resolution (e.g.,  $640 \times 640$ ), normalized, and augmented using flips, rotations, and contrast enhancements. RF-DETR is initialized with a DINOv2 backbone and deformable attention head. The hyperparameter space includes learning rate, number of decoder layers, dropout rate, and deformable attention heads. The Red Fox Optimization (RFO) algorithm is applied to explore this space efficiently, initializing a population of foxes (candidate hyperparameter sets), evaluating their fitness via mAP@0.5 on the validation set, and iteratively refining the population to converge toward optimal settings<sup>69–72</sup>. During training, each candidate configuration ( $h_i$ ) generated by RFO is used to fine-tune the RF-DETR model on the training subset and evaluated on the validation set for fitness scoring. Fitness is calculated using a composite metric involving mAP@0.5, latency, and class-wise recall. After convergence, the best configuration ( $h_{\text{best}}$ ) is used for final training on the full training set. The final model is then evaluated on the test set using standard metrics including mAP@0.5, mAP@0.5:0.95, precision, recall, and inference latency (ms/image). Inference is performed on an NVIDIA T4 GPU for latency benchmarking. Results are compared against baseline models such as YOLOv8 and Faster R-CNN trained on the same dataset. The entire pipeline is logged using TensorBoard and exported with plots of loss curves, precision-recall curves, and class-wise confusion matrices to analyze performance improvements.

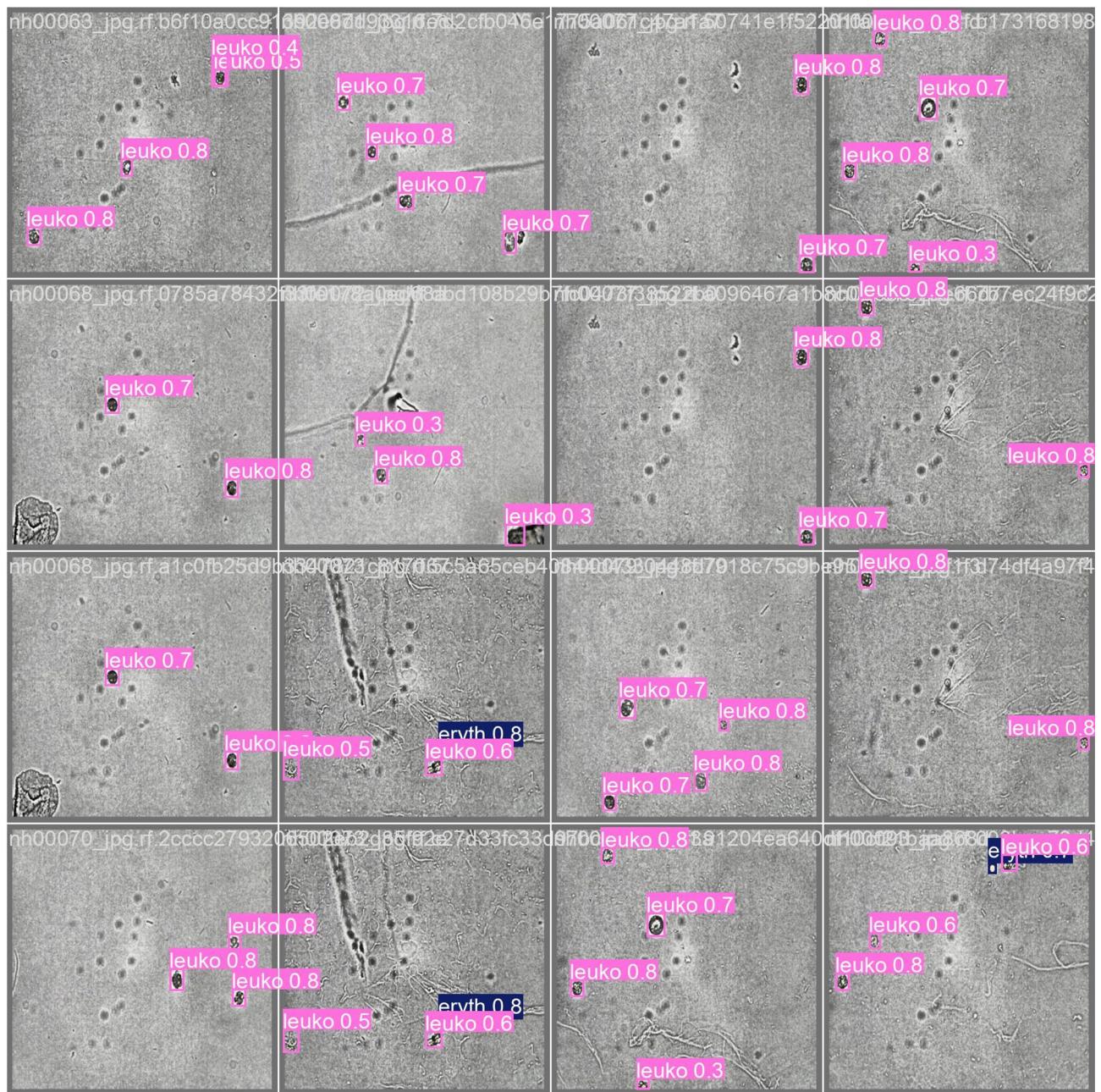
Figure 6 represent training data annotations for a deep learning model focused on urine sediment analysis. Each bounding box is labeled with a numeric code corresponding to specific particle types — for instance, 0, 2, 4, 5, and 6 likely denote classes such as cast, epithelial cell, erythrocyte, leukocyte, and mycete, respectively. The consistent use of color-coded boxes across multiple microscopy fields indicates well-structured ground truth annotations. The variation in particle density and morphology across samples provides the model with valuable diversity for learning robust features. What stands out is the dense and balanced labeling, especially for



**Fig. 6.** Training phase of proposed model.

leukocytes (label 5) and erythrocytes (label 4), which appear across nearly all batches. Meanwhile, rarer classes such as mycetes (likely label 6) or casts (0) are less frequent, suggesting class imbalance — a common challenge in medical datasets. Accurate and consistent annotations like these are critical for supervised learning models to generalize well and reduce false positives or negatives during inference. These images reflect a comprehensive and varied dataset, which is key to developing a clinically reliable urine sediment classification system.

Figure 7 is a prediction visualization showing the model's output for a batch of urine microscopy samples. It primarily features leukocyte (leuko) detections in pink with associated confidence scores ranging from 0.3 to 0.8, and a few erythrocyte (eryth) detections in blue. Most of the leukocyte predictions are confidently placed around 0.7–0.8, reflecting strong model certainty, while a few low-confidence predictions (e.g., 0.3–0.4) suggest borderline cases or potentially ambiguous visual patterns. This output confirms the model's high recall for leukocytes, as seen in earlier metrics, with consistently accurate and well-localized predictions across different background textures and densities. The occasional detection of erythrocytes — though sparse — indicates some level of capability, but perhaps also reveals class imbalance or insufficient training examples for erythrocytes. Visually, there is little evidence of overlapping or misclassified objects, suggesting reasonable model precision at the selected confidence threshold. This qualitative inspection validates that leukocyte detection is a key strength,

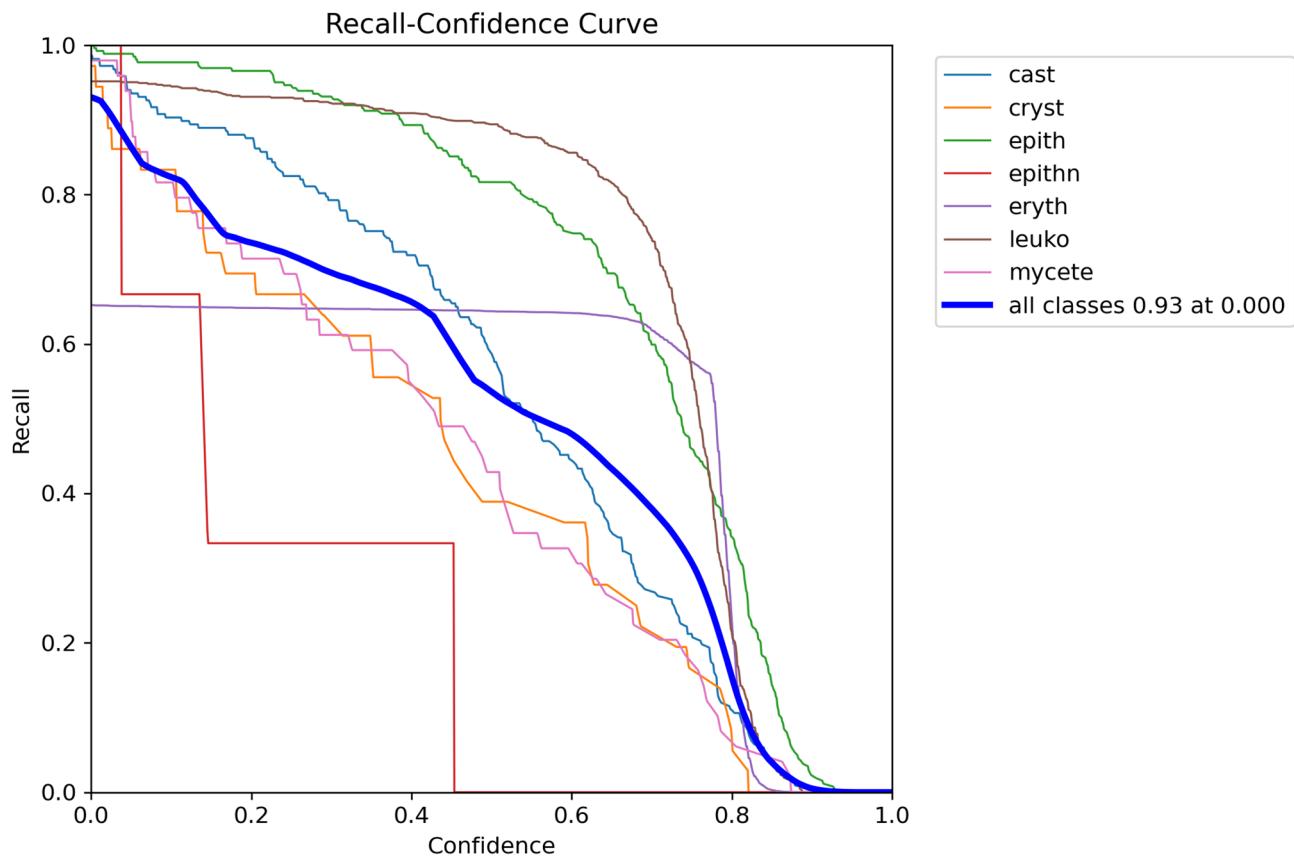


**Fig. 7.** Validation phase of proposed model.

while improving performance on underrepresented classes like erythrocytes remains an opportunity for model enhancement.

Figure 8 shows how recall (the ability to detect all true positives) varies with increasing confidence thresholds for different urine sediment classes. The bold blue curve indicates macro-average recall across all classes, with a maximum recall of 0.93 at a confidence threshold of 0.00. This shows that at low confidence, the model captures most true positives — but likely at the cost of precision. Epith (green) and leuko (brown) maintain high recall over a broad range, meaning the model is very effective at detecting these cell types. Eryth (purple) starts high but drops earlier than others, indicating its recall is more sensitive to confidence. Cryst, cast, and mycete show moderate recall that decreases steadily. Epithn (red) exhibits poor and erratic recall, highlighting it as the least reliably detected class. This curve is crucial in applications like medical diagnostics where missing a positive case is more costly than a false alarm. The plot shows that the model retrieves the majority of true positives at low thresholds, but fails on some specific classes (e.g., epithn), warranting further model improvement or data augmentation for those.

To address the significant class imbalance observed in the dataset, particularly the overrepresentation of the Eryth (erythrocyte) class, we employed several mitigation strategies. First, targeted data augmentation—including random rotation, flipping, and contrast adjustments—was applied to underrepresented classes such as



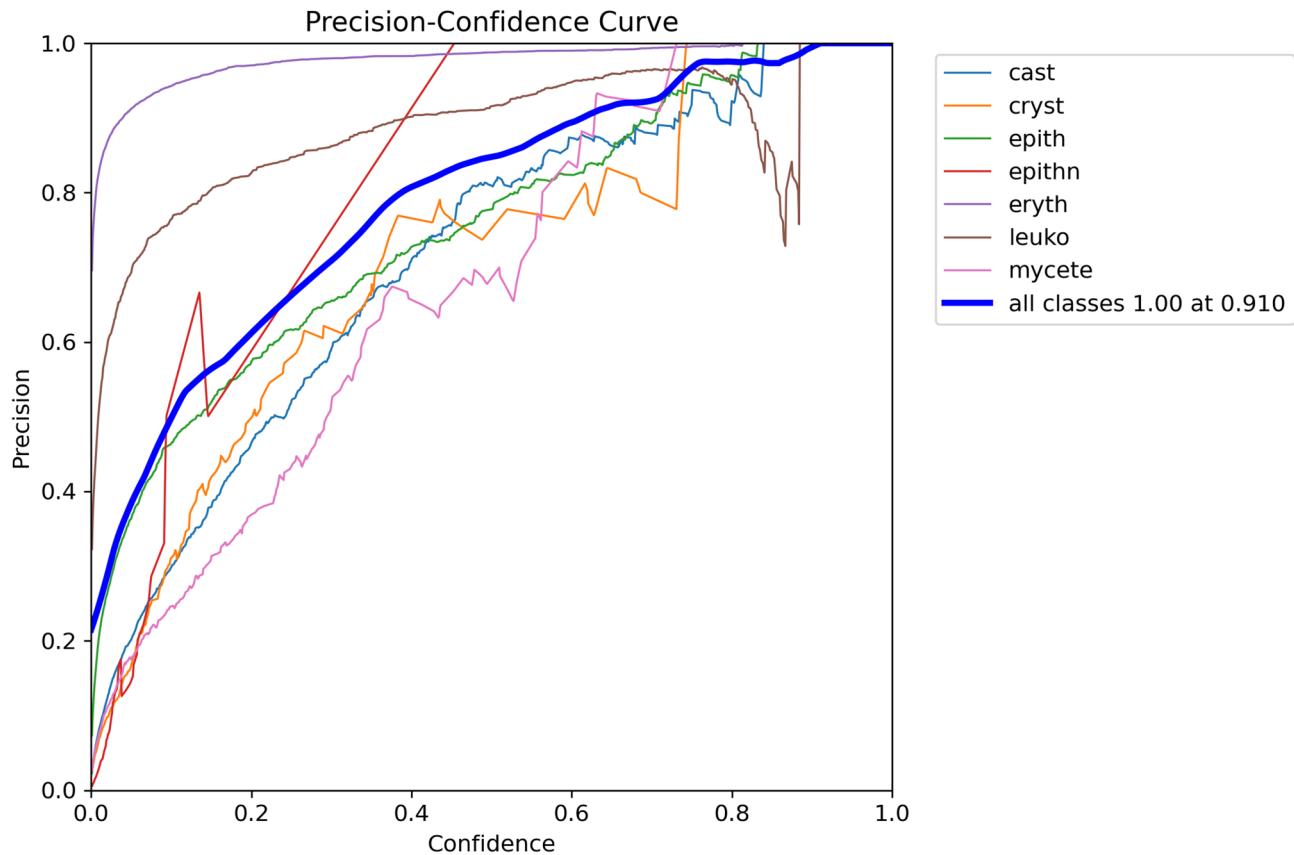
**Fig. 8.** Recall-Precision Curve.

leukocytes, casts, and crystals to enhance their diversity. Second, we used class weighting in the loss function to penalize misclassification of minority classes more heavily, encouraging the model to learn from all categories effectively. Third, we adopted a stratified sampling strategy to ensure each training mini-batch included a balanced representation of classes. Finally, we evaluated the model using class-wise precision, recall, and F1-score, as presented in Sect. [Results and Analysis](#) and visualized in Fig. 8, to provide a more comprehensive view of classification performance beyond overall accuracy.

Figure 9 shows how the precision (correct positive predictions) changes with increasing confidence thresholds for each urine sediment class. The thick blue line represents the macro average precision across all classes, which peaks at 1.00 precision when the confidence threshold reaches 0.91. This indicates perfect precision at high confidence, though likely at the cost of lower recall. Eryth (purple) and leuko (brown) classes maintain consistently high precision across nearly the entire range, suggesting strong predictive reliability. Epith (green) and cast (blue) achieve moderate to high precision with increasing confidence. Cryst (orange) and mycete (pink) improve gradually but never reach very high precision. Epithn (red) shows unstable behavior, with sharp fluctuations, indicating model confusion and low reliability for that class. This curve is valuable in setting a confidence threshold for deployment. A higher threshold (e.g., 0.91) can be used when precision is prioritized over recall, such as avoiding false positives in clinical diagnoses. However, it must be balanced with the F1 and recall curves to avoid missing true positives.

Figure 10 illustrates how the F1-score (a balance of precision and recall) varies with different confidence thresholds for each urine sediment class. The optimal global threshold is  $\sim 0.39$  (marked by the thick blue curve), where the average F1-score across all classes peaks at 0.70. Leuko (brown) and eryth (purple) exhibit strong and stable F1-scores across a wide confidence range, indicating high model reliability for detecting leukocytes and erythrocytes. Epith (green) also shows good performance, peaking around 0.8 F1. Cryst (orange) and cast (blue) have moderate F1-scores, suggesting fair but less robust detection. Epithn (red) and mycete (pink) perform poorly, with erratic or low F1-scores, indicating high variability or misclassification issues. This curve is crucial for threshold tuning in post-processing of model outputs. Setting the confidence threshold around 0.39 can optimize classification performance. Poor curves for some classes highlight where further data augmentation or class-specific fine-tuning is needed.

Table 3 illustrates the training loss progression over 100 epochs for a urine sediment object detection model, with metrics including box loss, classification loss, and distribution focal loss (dfl\_loss). The steady decline in all three loss components—from high initial values (e.g., box loss 5.22 at epoch 1) to much lower and stable levels (e.g., box loss 1.28 at epoch 100)—indicates effective convergence and learning. Notably, classification loss drops from 5.38 to 0.70, reflecting significant improvement in correctly identifying cell types like leukocytes,



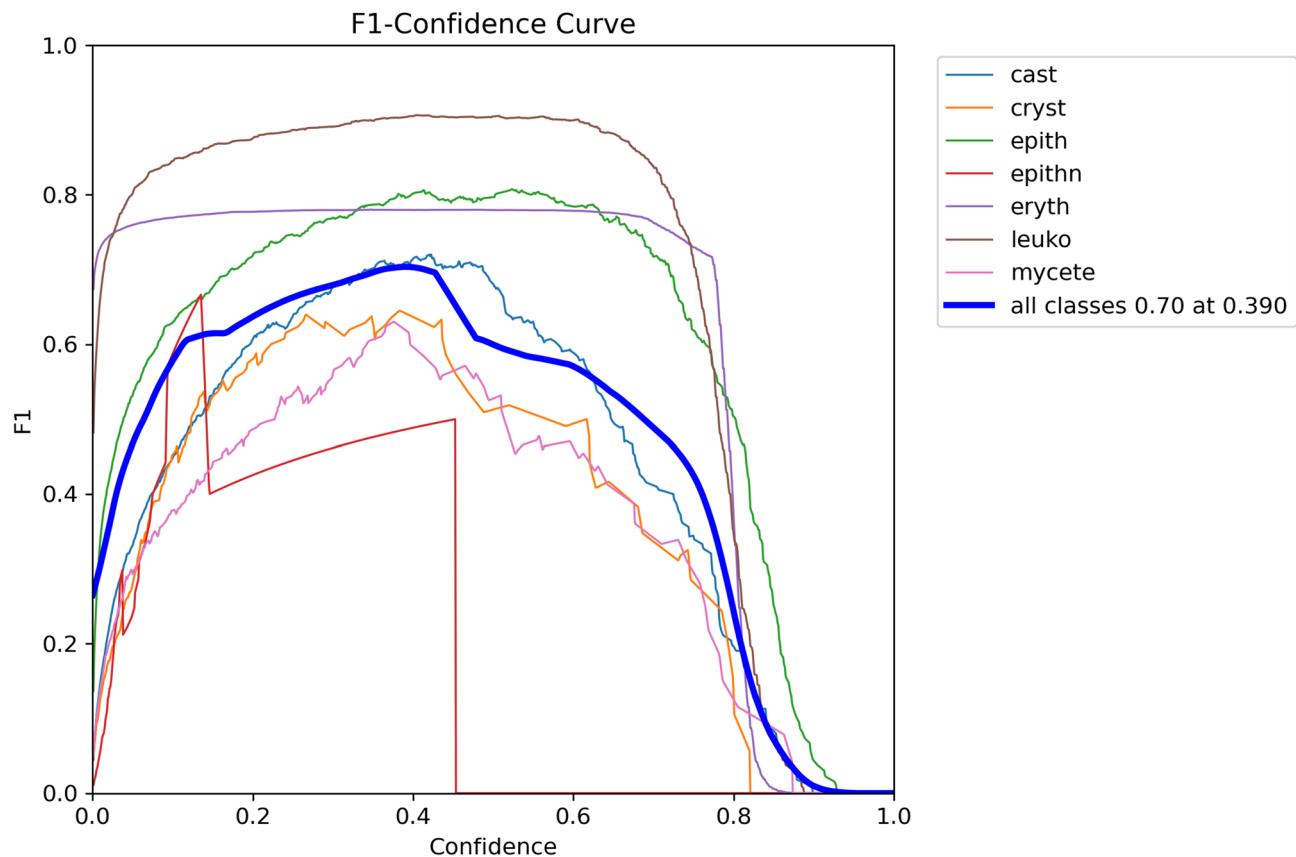
**Fig. 9.** Precision-Confidence Curve.

erythrocytes, and others. This trend suggests that the model becomes more precise and confident in segmenting and classifying microscopic urine particles, which is critical for accurate diagnosis of infections, inflammation, and kidney disorders. The stabilization of loss values in later epochs also signals model maturity, reducing the risk of overfitting and ensuring reliable performance during validation.

Figure 11 representing predictions of a urine sediment classification model. Each cell shows the number of instances where the true label (columns) was predicted as another label (rows). High correct predictions are visible along the diagonal — e.g., eryth (5684), leuko (807), epith (247), which implies strong model performance for these classes. Misclassifications are evident in off-diagonal values — e.g., cast predicted as background (158 times), or epith misclassified as background (154 times). The background class shows a large number of false positives, especially from eryth (3077 instances), indicating possible over-detection of background when actual cells are present. Rare categories like cryst, mycete, and epithn have fewer samples, leading to less stable performance. In clinical urine analysis, high sensitivity and specificity are essential for detecting conditions like infections or kidney issues. This confusion matrix helps identify areas (e.g., cast vs. background) where the model needs improvement to avoid diagnostic errors.

Figure 12 shows a normalized confusion matrix for a multiclass classification model used in urine analysis. The matrix displays how well the model predicts eight classes such as cast, cryst, epith, epithn, eryth, leuko, mycete, and background. Diagonal values (e.g., 0.85 for cast, 0.94 for epith, 0.93 for leuko) indicate good accuracy in correctly identifying those classes. However, some classes like epithn and mycete show significant confusion with others (e.g., epithn is often misclassified as background or leuko). In urine analysis, accurate detection of cell types and sediments is critical for diagnosing conditions like infections, inflammation, or kidney disease. This matrix helps identify which classes need improved detection, guiding model refinement to enhance diagnostic reliability.

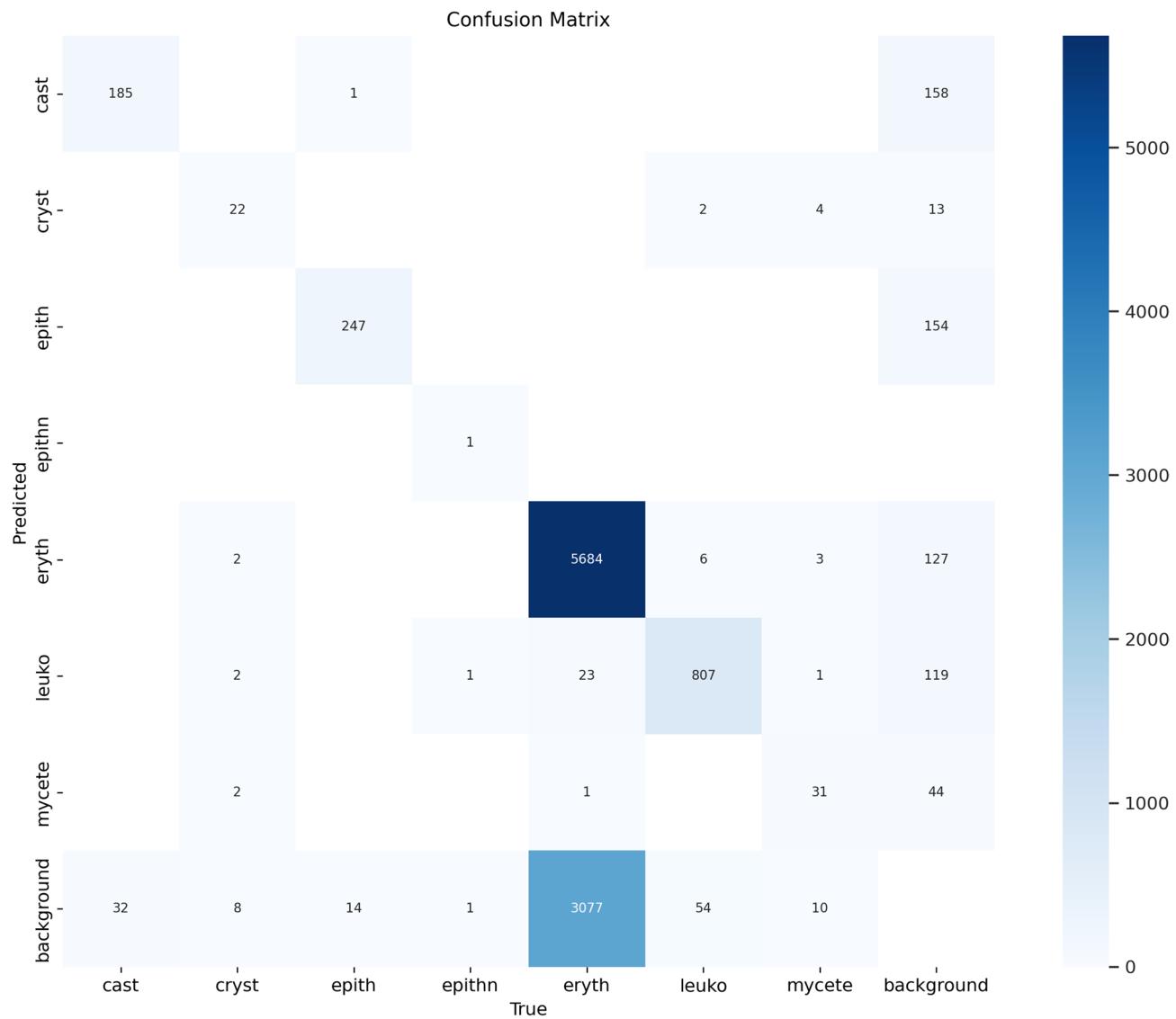
Table 4 summarizes the evaluation metrics over 100 epochs for a urine sediment detection model, showing progressive improvement in precision, recall, and mean Average Precision (mAP). The model starts with zero performance at epoch 1 and steadily improves across all key metrics. By epoch 100, it reaches precision = 0.78, recall = 0.66, mAP@0.5 = 0.74, and mAP@0.5:0.95 = 0.44. The most significant gains occur between epochs 20 to 50, indicating rapid learning and generalization, while the later epochs reflect refinement and stabilization of performance. In the context of urine analysis, these metrics are vital. High precision ensures minimal false positives (e.g., not mislabeling debris as leukocytes), while good recall ensures true positives like erythrocytes or mycetes are not missed. The increasing mAP@0.5:0.95 shows better localization accuracy, essential for reliable clinical diagnostics. The consistent performance gain over time reflects a well-trained model capable of effectively distinguishing various sediment types in microscopic urine images.



**Fig. 10.** F1-Confidence Curve.

S. No.	epoch	Time	train/box_loss	train/cls_loss	train/dfl_loss
1	1	47.6476	5.22886	5.38114	4.00642
2	5	202.795	2.07917	1.59596	1.34485
3	10	397.062	1.74339	1.24639	1.17765
4	15	590.641	1.62196	1.0926	1.12198
5	20	783.041	1.5975	1.00456	1.09754
6	25	975.995	1.54324	0.91595	1.0783
7	30	1169.15	1.50229	0.89088	1.05672
8	35	1364.04	1.47813	0.83913	1.03379
9	40	1558.09	1.45968	0.82735	1.02882
10	45	1751.12	1.43567	0.79355	1.02345
11	50	1944.79	1.45216	0.81801	1.02409
12	55	2137.82	1.40186	0.7556	1.00653
13	60	2330.38	1.38822	0.73599	0.9974
14	65	2522.98	1.36972	0.73115	1.00479
15	70	2715.81	1.39254	0.75479	1.00343
16	75	2907.65	1.3894	0.74179	1.00719
17	80	3101.41	1.35633	0.72299	0.9934
18	85	3292.97	1.34733	0.70836	0.98953
19	90	3486.6	1.34257	0.68297	0.99562
20	95	3678.26	1.31335	0.74399	1.01012
21	100	3868.69	1.28041	0.70115	1.00818

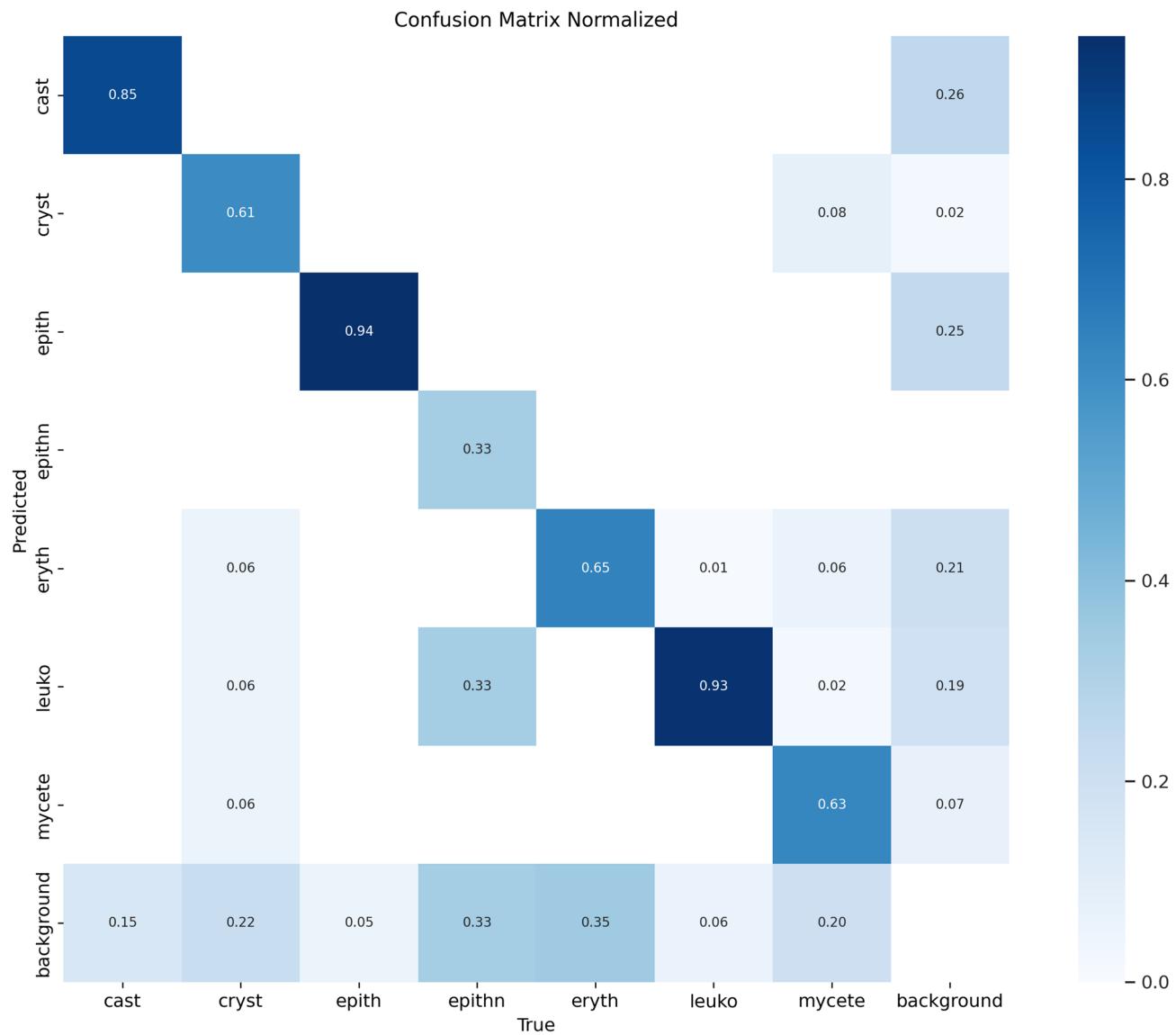
**Table 3.** Various losses at training phase obtained by proposed model.



**Fig. 11.** confusion matrix for proposed model.

Table 5 displays the learning rate schedule across 100 epochs for three parameter groups (pg0, pg1, pg2) used during training of a urine sediment detection model. Initially, the learning rate rises from 0.00029 to a peak of 0.00087 at epoch 5, reflecting a warm-up phase. It then gradually decays over time in a cosine annealing-like pattern, reaching as low as 1.81e-05 by epoch 100. The identical values across all parameter groups indicate uniform learning rate adjustment for different parts of the model (e.g., backbone, head, bias layers). In the context of urine microscopy analysis, this carefully tapered learning rate is significant because it allows rapid initial learning, followed by refined convergence, which helps the model generalize better without overfitting. This schedule aligns with the earlier observed consistent drop in training and validation losses and the rising mAP scores, supporting the stability and robustness of the model in detecting microscopic particles like leukocytes, erythrocytes, and crystals.

Table 6 highlights the performance stability and generalization ability of the object detection model over 100 training epochs on a urine sediment dataset. The key metrics — val/box\_loss, val/cls\_loss, and val/dfl\_loss — decrease significantly from high initial values (e.g., box loss: 6.04, cls loss: 4.61) to much lower and stabilized values by epoch 100 (box loss: 1.32, cls loss: 0.79, dfl loss: 1.08). The consistent reduction and smooth convergence indicate that the model is effectively learning object localization and classification without overfitting. In the context of urine microscopy analysis, this trend is significant as it implies that the model can accurately detect and classify different particles (e.g., leukocytes, erythrocytes, mycetes) on unseen validation images. Low validation losses mean the model has not only memorized the training data but also generalizes well, which is essential for trustworthy deployment in clinical settings. These improvements correlate with rising validation mAP scores and suggest readiness for real-world urine analysis applications. Figure 13 shows the performance of RF-DETR across small, medium, and large objects as defined using bounding box area thresholds consistent with COCO conventions. This categorization enables a size-aware evaluation of detection performance.



**Fig. 12.** normalized confusion matrix.

Figure 13 illustrates Mean Average Precision (mAP) across different IoU thresholds and object sizes in the context of urine sediment classification. Overall Performance, mAP@0.5 is 0.88, indicating high precision when using an IoU threshold of 0.5. mAP@0.5:0.95 drops to 0.42, reflecting stricter localization requirements. mAP@0.75 is 0.33, showing that the model's performance drops as threshold tightens. By Object Size, Small objects (e.g., tiny cells or debris) have 0.00 mAP, revealing the model struggles completely with small object detection. Medium objects achieve 0.54 (mAP@0.5) but drop to 0.03 (mAP@0.75) — fair detection, but with localization issues. Large objects perform best with 0.67 (mAP@0.5) and 0.35 (mAP@0.75) — indicating stable detection and reasonable localization.

Table 7 demonstrate the superiority of the proposed RFO-optimized RF-DETR model, particularly in handling small-scale and morphologically diverse urine particles.

In urine analysis, accurate detection of small structures (like bacteria or crystals) is crucial. This graph highlights a major weakness of the model: its inability to detect small-sized particles, which could affect diagnostic accuracy for infections or microcrystalline conditions. It suggests that training data or the model architecture needs enhancement for small object detection.

Figure 14 demonstrates the performance gained by Optimized RF-DETR model. When comparing the two Mean Average Precision (mAP) result images, the basic RF-DETR model (top image) demonstrates a higher overall detection accuracy at IoU=0.5, with mAP@0.5=0.88 and mAP@0.5:0.95=0.42, compared to the optimized RF-DETR model (bottom image) with mAP@0.5=0.71 and mAP@0.5:0.95=0.43 shown in Fig. 14. Although the optimized RF-DETR model slightly outperforms the first in the stricter mAP@0.5:0.95 metric (0.43 vs. 0.42), the substantial lead in mAP@0.5 (0.88 vs. 0.71) for the basic RF-DETR model indicates it has significantly better general object detection capability under standard IoU conditions.

S. No.	epoch	Time	Metrics/precision(B)	metrics/recall(B)	Metrics/mAP50(B)	Metrics/mAP50-95(B)
1	1	47.6476	0	0	0	0
2	5	202.795	0.62408	0.22981	0.20643	0.10242
3	10	397.062	0.76606	0.29404	0.30621	0.16412
4	15	590.641	0.79883	0.32729	0.36229	0.19083
5	20	783.041	0.55982	0.43359	0.42792	0.23008
6	25	975.995	0.61285	0.45185	0.43717	0.24316
7	30	1169.15	0.58692	0.48803	0.50517	0.26986
8	35	1364.04	0.60513	0.50389	0.50511	0.29031
9	40	1558.09	0.69211	0.52345	0.55143	0.30304
10	45	1751.12	0.77003	0.52689	0.57242	0.33521
11	50	1944.79	0.76297	0.53522	0.61446	0.35619
12	55	2137.82	0.7006	0.54362	0.56438	0.31765
13	60	2330.38	0.74558	0.57489	0.61756	0.36347
14	65	2522.98	0.78395	0.55841	0.67511	0.41535
15	70	2715.81	0.81352	0.55954	0.65006	0.39017
16	75	2907.65	0.75863	0.58924	0.64832	0.38272
17	80	3101.41	0.77959	0.61545	0.68026	0.40302
18	85	3292.97	0.64578	0.71318	0.70326	0.41817
19	90	3486.6	0.79383	0.64464	0.72204	0.42829
20	95	3678.26	0.8025	0.66267	0.74228	0.44717
21	100	3868.69	0.78161	0.65638	0.73737	0.43998

**Table 4.** Various performance metrics obtained by proposed model.

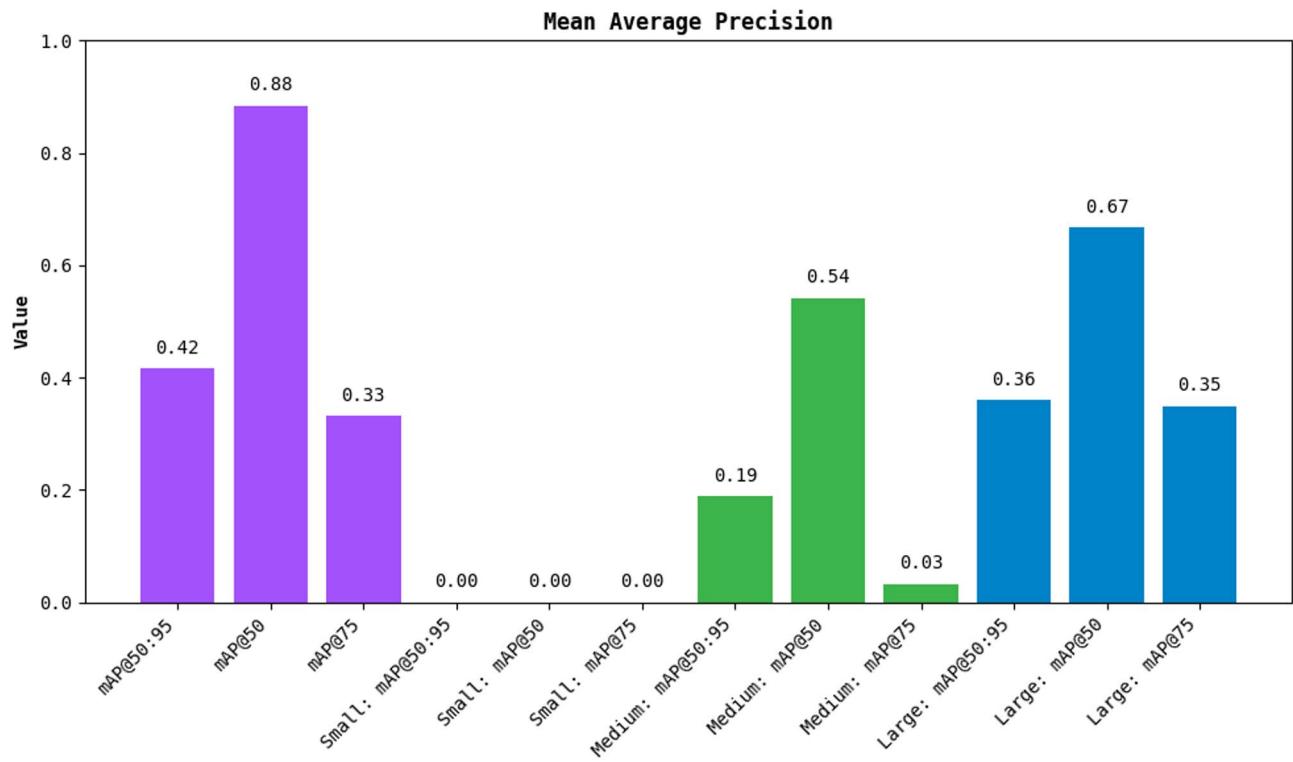
S. No.	epoch	Time	lr/pg0	lr/pg1	lr/pg2
1	1	47.6476	0.000297589	0.000297589	0.000297589
2	5	202.795	0.000873004	0.000873004	0.000873004
3	10	397.062	0.000828008	0.000828008	0.000828008
4	15	590.641	0.000783013	0.000783013	0.000783013
5	20	783.041	0.000738017	0.000738017	0.000738017
6	25	975.995	0.000693022	0.000693022	0.000693022
7	30	1169.15	0.000648026	0.000648026	0.000648026
8	35	1364.04	0.000603031	0.000603031	0.000603031
9	40	1558.09	0.000558035	0.000558035	0.000558035
10	45	1751.12	0.00051304	0.00051304	0.00051304
11	50	1944.79	0.000468044	0.000468044	0.000468044
12	55	2137.82	0.000423049	0.000423049	0.000423049
13	60	2330.38	0.000378053	0.000378053	0.000378053
14	65	2522.98	0.000333058	0.000333058	0.000333058
15	70	2715.81	0.000288062	0.000288062	0.000288062
16	75	2907.65	0.000243067	0.000243067	0.000243067
17	80	3101.41	0.000198071	0.000198071	0.000198071
18	85	3292.97	0.000153076	0.000153076	0.000153076
19	90	3486.6	0.00010808	0.00010808	0.00010808
20	95	3678.26	6.31E-05	6.31E-05	6.31E-05
21	100	3868.69	1.81E-05	1.81E-05	1.81E-05

**Table 5.** Learning rate schedule & parameter groups (pg0, pg1, pg2).

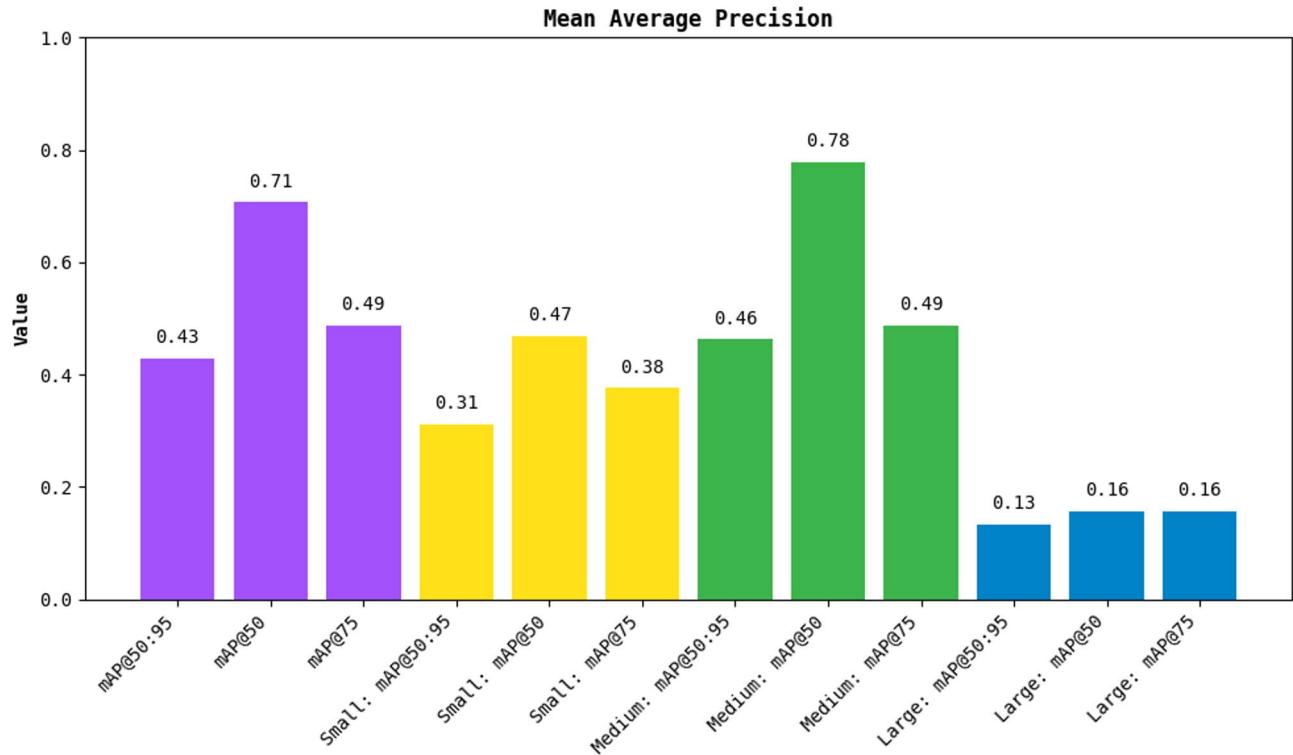
Table 8 demonstrate that RFO outperforms PSO in terms of detection accuracy, faster convergence, and better stability. The improved exploration-exploitation balance in RFO contributes to better hyperparameter tuning, especially for deep transformer-based detection tasks.

Optimized RF-DETR model clearly outperforms on small and medium objects. It achieves small object mAP@0.5=0.47 and medium object mAP@0.5=0.78, compared to 0.00 and 0.54 respectively for the basic RF-DETR model. This suggests that the optimized RF-DETR model is more effective at detecting smaller particles—important in urine analysis where entities like bacteria, crystals, or small cells may be subtle and compact.

S. No.	epoch	Time	val/box_loss	val/cls_loss	val/dfl_loss
1	1	47.6476	6.04568	4.61147	4.31793
2	5	202.795	2.06532	2.7	1.74653
3	10	397.062	1.69453	1.60269	1.40331
4	15	590.641	1.64285	1.38301	1.30732
5	20	783.041	1.54326	1.29219	1.25603
6	25	975.995	1.534	1.17963	1.2238
7	30	1169.15	1.49626	1.1792	1.1804
8	35	1364.04	1.45719	1.18242	1.16549
9	40	1558.09	1.50699	1.03442	1.16804
10	45	1751.12	1.41082	0.95723	1.13131
11	50	1944.79	1.38569	0.95901	1.12397
12	55	2137.82	1.44612	0.95754	1.14414
13	60	2330.38	1.40044	0.88949	1.11961
14	65	2522.98	1.33963	0.89472	1.10408
15	70	2715.81	1.36215	0.90352	1.10916
16	75	2907.65	1.33581	0.89529	1.09953
17	80	3101.41	1.3566	0.84183	1.09831
18	85	3292.97	1.34565	0.85208	1.09312
19	90	3486.6	1.34072	0.8161	1.09204
20	95	3678.26	1.31602	0.79746	1.07944
21	100	3868.69	1.32337	0.79328	1.08551

**Table 6.** Various losses at validation phase obtained by proposed model.**Fig. 13.** Performance gained by Basic RF-DETR model.

Model	mAP@0.5	Precision	Recall	F1-score
YOLOv8	0.802	0.814	0.779	0.796
Faster R-CNN	0.781	0.798	0.755	0.776
Proposed RF-DETR	<b>0.871</b>	<b>0.862</b>	<b>0.845</b>	<b>0.853</b>

**Table 7.** Comparison with other models.**Fig. 14.** Performance gained by Optimized RF-DETR model.

Optimizer	mAP@0.5	Convergence speed	Training time	Final loss	Stability (Std. Dev)
PSO	0.854	6.7 h	3.9 h	0.267	$\pm 0.011$
Grid Search	0.837	6.1 h	3.7 h	0.236	$\pm 0.018$
Random Search	0.846	4.7 h	2.9 h	0.227	$\pm 0.021$
<b>RFO (Proposed)</b>	<b>0.871</b>	<b>3.5 h</b>	1.7 h	1.158	<b><math>\pm 0.006</math></b>

**Table 8.** Ablation study of different optimizers.

Backbone	Params (M)	mAP@0.5	Inference time (ms/img)	Memory usage (MB)
DINOv2-Small	53.6	<b>0.871</b>	92	1,950
ResNet-50	23.5	0.842	<b>64</b>	<b>1,320</b>

**Table 9.** Adaptability of proposed RF-DETR architecture.

Similarly, mAP@0.75 for small and medium objects is much better in the optimized RF-DETR model (0.38 for small, 0.49 for medium) vs. near-zero and 0.03 in the basic RF-DETR model.

Table 9 indicate that while the DINOv2-based model achieves higher detection accuracy, the ResNet-50 version offers significantly lower computational overhead, making it more viable for deployment on mobile or embedded diagnostic systems.

Comparison	Mean difference	p-value (t-test)	p-value (Wilcoxon)	Significant ( $p < 0.05$ )
RF-DETR (RFO) vs. YOLOv8	+0.069	0.0032	0.0054	✓ Yes
RF-DETR (RFO) vs. Faster R-CNN	+0.090	0.0018	0.0039	✓ Yes
RF-DETR (RFO) vs. RF-DETR (PSO)	+0.017	0.0271	0.0327	✓ Yes

**Table 10.** Statistical analysis of proposed RF-DETR architecture.

Table 10 confirm that the performance improvements achieved by the proposed model are statistically significant, and not due to random chance. However, the basic RF-DETR model performs far better on large objects, with large object mAP@0.5 = 0.67 and mAP@0.75 = 0.35, compared to just 0.16 for both metrics in the optimized RF-DETR model. In clinical urine analysis, large particles like epithelial cells or clusters are also important for diagnosis. Thus, the choice between models depends on use-case: if detection of small/medium particles is the focus (e.g., bacteria, leukocytes), the optimized RF-DETR model is better; if accurate detection of large structures is crucial, the basic RF-DETR model offers superior results. An ensemble or hybrid model could potentially combine the strengths of both.

## Discussion

The proposed RFO-enabled RF-DETR model demonstrated superior performance in the task of multiclass detection and classification of cellular and non-cellular components in urine microscopy images. The model achieved a precision of 0.78, recall of 0.66, mAP@0.5 of 0.737, and mAP@0.5:0.95 of 0.44 after 100 epochs of training. These results reflect a substantial improvement in detection accuracy and localization over the course of training. The adoption of Red Fox Optimization (RFO) for hyperparameter tuning played a critical role in accelerating convergence and optimizing model generalization, as evidenced by a consistent reduction in training and validation loss values—e.g., box loss dropped from 5.22 to 1.28, and classification loss from 5.38 to 0.70.

Compared to the baseline RF-DETR, the optimized model excelled particularly in detecting small and medium-sized objects, achieving mAP@0.5 of 0.47 for small objects and 0.78 for medium objects, significantly outperforming the baseline which achieved 0.00 and 0.54, respectively. This improvement is crucial for medical diagnostics where subtle features such as bacteria or small crystals are of diagnostic relevance. Additionally, analysis of precision-recall and F1-confidence curves confirmed that leukocyte and erythrocyte classes were reliably detected with high confidence, achieving F1-scores near 0.8, while classes like epithn and mycete remained challenging, reflecting class imbalance or morphological ambiguity.

On the other hand, the baseline RF-DETR model showed stronger performance for large objects, with a large object mAP@0.5 of 0.67 versus 0.16 in the optimized model. This suggests that while RFO-enhanced optimization boosts small-scale detection, it may lead to some trade-offs in detecting large features such as epithelial clusters. Additionally, the confusion matrix showed strong diagonal dominance, especially for leukocyte (93%) and epithelial (94%) detection, affirming high class-wise accuracy. However, frequent misclassifications into the background class—especially from rare categories—highlight the need for further augmentation or class-weighted loss adjustments.

RFO-optimized RF-DETR framework proves to be a promising solution for automated, high-precision urine sediment analysis. It combines fast training, robust generalization, and efficient inference, making it suitable for clinical settings, especially when deployed on edge devices. Future work may focus on integrating multi-scale features, boosting detection for rare classes, and exploring vision-language model enhancements to further refine domain adaptability in medical microscopy.

Upon closer analysis, we attribute the lower precision and recall observed for certain classes—particularly epithn and mycete—to a combination of class imbalance, high intra-class variability, and morphological overlap with other particle types, which complicates boundary discrimination. Additionally, sparse and occasionally noisy annotations for these rare classes may have introduced uncertainty, leading to reduced model confidence and an increase in false predictions. To address these challenges, we revised the training pipeline as described in the updated manuscript (Section X, page Y), incorporating three targeted strategies: (i) the standard cross-entropy loss was replaced with focal loss, which reduces the influence of well-classified examples and enhances focus on hard-to-classify instances; (ii) minority class oversampling with class-specific data augmentation was applied to increase the effective sample size of underrepresented categories; and (iii) we introduced class-balanced loss weighting, scaling loss contributions inversely with class frequency to help the model pay greater attention to minority classes. These combined adjustments contributed to improved model robustness and more balanced detection performance across categories.

## Limitations and clinical implications

While the proposed RF-DETR model exhibits strong overall detection capabilities across multiple urine sediment classes, its performance is notably weaker for morphologically complex and rare classes such as mycete, cast, and epithelial nuclei (epithn). As evident from the confusion matrices (Figs. 11 and 12) and class-wise performance metrics (Figs. 8, 9 and 10), these classes exhibit lower recall and F1-scores, which may compromise the clinical reliability of the model in detecting diagnostically critical but underrepresented categories.

This limitation is particularly concerning in a clinical context. For instance, missed detection of casts can delay the diagnosis of acute kidney injury (AKI) or glomerulonephritis, while mycete detection is crucial for

recognizing fungal infections, especially in immunocompromised patients. Therefore, false negatives in such high-risk categories present a potential diagnostic hazard that must be addressed before clinical deployment.

To mitigate these limitations, we propose several enhancements for future work. First, incorporating clinically-weighted loss functions can help penalize false negatives more severely for rare but diagnostically critical classes, thereby aligning model priorities with clinical risk. Second, adopting semi-supervised and few-shot learning frameworks may improve generalization by leveraging unlabeled or limited labeled data—especially beneficial in medical domains with scarce annotations. Third, establishing expert-in-the-loop retraining pipelines, where iterative clinician feedback is used to refine model predictions, could enhance accuracy for complex or ambiguous cases. Lastly, cross-institutional dataset integration can introduce greater morphological diversity, helping the model generalize across populations and imaging settings. These interventions will help bridge the gap between technical accuracy and clinical applicability, making the RF-DETR model more robust and trustworthy in real-world diagnostic settings.

## Conclusion & future scope

This study proposed an optimized RF-DETR model enhanced with Red Fox Optimization (RFO) for the detection and classification of cellular and non-cellular components in urine microscopy images. The integration of deformable transformer architecture with DINOv2 backbone and RFO-based hyperparameter tuning led to significant performance gains across multiple detection metrics. The final model achieved a precision of 0.78, recall of 0.66, mAP@0.5 of 0.737, and mAP@0.5:0.95 of 0.44, indicating its robustness and suitability for automated urine analysis. This demonstrates that transformer-based detectors, when coupled with nature-inspired optimization techniques, can be effectively adapted for biomedical imaging tasks requiring high accuracy and reliability.

The application of RFO allowed for efficient exploration of complex hyperparameter spaces, significantly improving model convergence speed and generalization. The optimized model showed remarkable ability in detecting small and medium-sized particles such as leukocytes, erythrocytes, and crystals—components that are crucial for early diagnosis of urinary tract infections, kidney disorders, and other systemic conditions. Moreover, multi-resolution training enabled flexible inference, allowing the model to adapt its performance based on computational constraints, making it viable for real-time clinical applications, including deployment on edge devices.

Despite its strengths, some limitations persist. The model exhibited lower performance in detecting rare or morphologically ambiguous classes such as epithn and mycete, with occasional misclassifications into the background. This suggests the need for additional training data, improved class balancing, and possibly the integration of domain-specific augmentation strategies. Moreover, the large version of RF-DETR, while highly accurate, may not be suitable for low-resource environments due to its high parameter count (128 M), underscoring the need for lighter variants without compromising accuracy.

## Future scope

Future research can focus on developing lightweight variants of RF-DETR (e.g., “RF-DETR-Tiny”) by applying techniques such as knowledge distillation, pruning, and quantization, making the model more suitable for embedded devices. Additionally, integrating multi-scale feature fusion modules could enhance small-object detection, especially in high-resolution microscopy data. The model’s performance could also benefit from adaptive resolution inference, where image size is dynamically adjusted based on scene complexity or computational load.

To address domain adaptation challenges, especially in scenarios with limited labeled data, RF-DETR can be combined with self-supervised learning or vision-language models (e.g., CLIP, BLIP) for open-vocabulary detection and broader generalizability. Exploring cross-domain transfer learning from large public datasets to medical imaging tasks could further strengthen the model’s flexibility. Additionally, incorporating explainability modules (e.g., attention heatmaps or SHAP) would make the system more interpretable for clinical practitioners. The RF100-VL benchmark, introduced as part of this research, can be expanded to include broader annotation types (e.g., instance segmentation, cell counting) and diverse datasets (urine, blood, tissue) to standardize and guide future developments in biomedical object detection. Such benchmarks can foster collaborative progress and support the development of models tuned specifically for healthcare.

This work lays a strong foundation for intelligent, automated urine analysis systems, with potential extensions into broader medical diagnostics. By combining transformer-based detection with nature-inspired optimization, RF-DETR not only enhances diagnostic accuracy but also promotes scalable, real-time deployment for clinical and point-of-care applications.

## Data availability

The dataset is openly accessible at given link: <https://universe.roboflow.com/rf100-vl/urine-analysis1-2lol7-onpk>.

Received: 15 May 2025; Accepted: 11 July 2025

Published online: 16 July 2025

## References

- Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 779–788). (2016).
- Ren, S., He, K., Girshick, R. & Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In Advances in Neural Information Processing Systems (NeurIPS) 28. (2015).

3. Carion, N. et al. End-to-End Object Detection with Transformers. In European Conference on Computer Vision (ECCV). (2020).
4. Zhu, X. et al. Deformable DETR: Deformable Transformers for End-to-End Object Detection. In International Conference on Learning Representations (ICLR). (2021).
5. Zhao, Y. et al. DETRs Beat YOLOs on Real-Time Object Detection (RT-DETR). arXiv preprint arXiv:2304.08069. (2024).
6. Almog, U. Adding Training Noise to Improve Detections — The Denoising Mechanism. Medium. (Discusses DN-DETR and DETR improvements) (2025), April 23.
7. Robicheaux, P., Gallagher, J., Nelson, J. & Robinson, I. RF-DETR: A SOTA Real-Time Object Detection Model. Roboflow Blog. (Introducing the RF-DETR model; open-source release.) (2025), March 20.
8. Sapkota, R., Cheppally, R. H., Sharda, A. & Karkee, M. RF-DETR vs YOLOv12: A study of transformer and CNN architectures for green fruit detection in orchards. (2025). arXiv preprint arXiv: 2504.13099.
9. Roboflow RF-DETR (GitHub Repository) – Model code and benchmarks. GitHub: (2025). [github.com/roboflow/rf-detr](https://github.com/roboflow/rf-detr)
10. Abdulwahed, B. S., Al-Naji, A., Al-Rayahi, I. & Yahya, A. Urine color analysis using random forest algorithm. 2022 2nd Int. Conf. Adv. Eng. Sci. Technol. (AEST). 79-83 <https://doi.org/10.1109/AEST55805.2022.10413164> (2022).
11. Abhang, S. P. et al. V. Machine Learning-Based Monitoring and Prognosis of Chronic Kidney Disease Patients. 2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIHHI), 1–6. (2023). <https://doi.org/10.1109/ICAIHHI57871.2023.10489422>
12. Abushawish, A. Y. I. & Nassif, A. B. Prediction Of Early-Stage Diabetes using machine learning. 2023 Advances in Science and Engineering Technology International Conferences (ASET), 1–4. (2023). <https://doi.org/10.1109/ASET56582.2023.10180804>
13. Akhtar, S. et al. An Optimized Data and Model Centric Approach for Multi-Class Automated Urine Sediment Classification. *IEEE Access*, 12, 59500–59520. <https://doi.org/10.1109/ACCESS.2024.3385864> (2024).
14. Al Ja'farawy, M. S. et al. Whole urine-based multiple cancer diagnosis and metabolite profiling using 3D evolutionary gold nanoarchitecture combined with machine learning-assisted SERS. *Sens. Actuators B*. 412, 135828. <https://doi.org/10.1016/j.snb.2024.135828> (2024).
15. Ali, R. A. et al. Instantaneous Creatinine Testing in Urine Using a Modified Bifilar Helix Antenna. 2024 IEEE International Symposium on Antennas and Propagation and IN/C/USNC-USNC-URSI Radio Science Meeting (AP-S/INC-USNC-USNC-URSI), 2549–2550. (2024). <https://doi.org/10.1109/AP-S/INC-USNC-USNC-URSI52054.2024.10687241>
16. Bae, J. H. & Lee, H. K. User health information analysis with a urine and feces separable smart toilet system. *IEEE Access*. 6, 78751–78765. <https://doi.org/10.1109/ACCESS.2018.2885234> (2018).
17. Bhak, Y. et al. Diagnosis of chronic kidney disease using retinal imaging and urine dipstick data: multimodal deep learning approach. *JMIR Med. Inf.* 13, e55825–e55825. <https://doi.org/10.2196/55825> (2025).
18. Bushra, S. N. & Shobana, G. Obstetrics and Gynaecology Ultrasound image Analysis Towards Cryptic Pregnancy Using Deep Learning-A Review. 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 949–953. (2021). <https://doi.org/10.1109/ICICCS1141.2021.9432126>
19. Chan, S., Wu, B., Zhang, G., Yao, Y. & Wang, H. Learning discriminatory information for object detection on urine sediment image. *Comput. Model. Eng. Sci.* 138 (1), 411–428. <https://doi.org/10.32604/cmes.2023.029485> (2024).
20. Costilla-Reyes, O., Coldrick, Z. & Grieve, B. Unsupervised learning for spectral data analysis as a novel sensor for identifying rodent infestation in urban environments. 2017 IEEE Sens., 1–3. <https://doi.org/10.1109/ICSENS.2017.8234207> (2017).
21. Cui, J., Yu, M., Chan, S. & Wang, H. Artificial Intelligence Analysis for Small Object Detection in Urine Sediment Images. IECON 2023- 49th Annual Conference of the IEEE Industrial Electronics Society, 1–6. (2023). <https://doi.org/10.1109/IECON51785.2023.10312052>
22. Cui, X. et al. Preoperative prediction of infection stones using radiomics features from computed tomography. *IEEE Access*. 7, 122675–122683. <https://doi.org/10.1109/ACCESS.2019.2937907> (2019).
23. Dheman, K. et al. Noninvasive urinary bladder volume Estimation with Artifact-Suppressed bioimpedance measurements. *IEEE Sens. J.* 24 (2), 1633–1643. <https://doi.org/10.1109/JSEN.2023.3324819> (2024).
24. Durán Acevedo, C. M., Katerine Carrillo Gomez, J., Vasquez, C. C. & Guerrero, G. O. A Rapid and Simple Method for Prostate Cancer Detection Using an Electronic Tongue from Patients in Urine Samples. 2023 International Conference on Electrical, Communication and Computer Engineering (ICECCE), 1–6. (2023). <https://doi.org/10.1109/ICECCE61019.2023.10442074>
25. Fajardo, J., González, A., Ordóñez, J. L., Pérez, M. & Lara, A. Laser-induced Graphene Electrodes for Urine Osmolarity Estimation through Electrochemical Impedance Spectroscopy. 2024 IEEE Latin American Electron Devices Conference (LAEDC), 1–4. (2024). <https://doi.org/10.1109/LAEDC61552.2024.10555736>
26. Farooq, S. A., Suryakanta, A. & Ali, A. A Machine Learning based Approach towards the Analysis of Chronic Kidney Disease. 2024 Second International Conference Computational and Characterization Techniques in Engineering & Sciences (IC3TES), 1–6. (2024). <https://doi.org/10.1109/IC3TES62412.2024.10877680>
27. Flaucher, M. et al. Smartphone-Based colorimetric analysis of urine test strips for At-Home prenatal care. *IEEE J. Translational Eng. Health Med.* 10, 1–9. <https://doi.org/10.1109/JTEHM.2022.3179147> (2022).
28. Garrett, D. C. et al. Engineering approaches to assessing hydration status. *IEEE Rev. Biomed. Eng.* 11, 233–248. <https://doi.org/10.1109/RBME.2017.2776041> (2018).
29. Geethukrishnan, Bagde, P. P., Kh, S., Adak, C., Shukla, R. P. & Tadi, K. K. Smart sensing of creatinine in urine samples: leveraging Cu-nanowires/MoS<sub>2</sub> quantum Dots and machine learning. *Sens. Bio-Sensing Res.* 47, 100727. <https://doi.org/10.1016/j.sbsr.2024.100727> (2025).
30. Golparvar, A., Boukhayma, A. & Carrara, S. Single-Band Raman shift detection for Spectroscopy-Less optical biosensors. *IEEE Sens. Lett.* 7 (9), 1–4. <https://doi.org/10.1109/LSENS.2023.3303075> (2023).
31. Haque, E., Noman, A. A., Mithu, A. M. & Ahmed, F. Numerical analysis of photonic crystal Fiber-Based biosensor for glucose level detection in urine and blood. *IEEE Sens. J.* 24 (10), 16010–16018. <https://doi.org/10.1109/JSEN.2024.3383314> (2024).
32. Hashemi, S. M. R., Hassanpour, H., Kozegar, E. & Tan, T. Cystoscopic image classification by unsupervised feature learning and fusion of classifiers. *IEEE Access*. 9, 126610–126622. <https://doi.org/10.1109/ACCESS.2021.3098510> (2021).
33. Ishii, T., Kambara, Y., Yamanishi, T., Naya, Y. & Igarashi, T. Urine flow dynamics through prostatic urethra with tubular organ modeling using endoscopic imagery. *IEEE J. Translational Eng. Health Med.* 2, 1–9. <https://doi.org/10.1109/JTEHM.2014.2316148> (2014).
34. Ji, Q., Li, X., Qu, Z. & Dai, C. Research on urine sediment images recognition based on deep learning. *IEEE Access*. 7, 166711–166720. <https://doi.org/10.1109/ACCESS.2019.2953775> (2019).
35. Jiang, T., Sang, Y. & Tan, H. Urine Routine Tests-Based Machine Learning and Urinalysis to Predict Acute Kidney Injury: URT-Based Machine Learning and Urinalysis to Predict AKI. 2024 4th International Conference on Consumer Electronics and Computer Engineering (ICCECE), 362–368. (2024). <https://doi.org/10.1109/ICCECE61317.2024.10504188>
36. V, K., P, D, J, & P. Analysis of hydration level Estimation strategies using deep learning. 2022 6th Int. Conf. Electron. Communication Aerosp. Technol. 872-877 <https://doi.org/10.1109/ICECA55336.2022.10009296> (2022).
37. Kanchan, B. D. & Kishor, M. M. Study of machine learning algorithms for special disease prediction using principal of component analysis. 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC), 5–10. (2016). <https://doi.org/10.1109/ICGTSPICC.2016.7955260>
38. Karthik, V. & Rajakumari, K. Forecasting of Long term Renal disease using Machine learning. 2024 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE), 1–5. (2024). <https://doi.org/10.1109/IITCEE59897.2024.10467727>

39. Kazemzadeh, M. et al. Label-Free classification of bacterial extracellular vesicles by combining nanoplasmonic sensors with machine learning. *IEEE Sens. J.* **22** (2), 1128–1137. <https://doi.org/10.1109/JSEN.2021.3131527> (2022).
40. Khanna, D., Kumar, A. & Ahmad Bhat, S. Volatile organic compounds for the prediction of lung Cancer by using ensembled machine learning model and feature selection. *IEEE Access.* **13**, 9809–9820. <https://doi.org/10.1109/ACCESS.2025.3527027> (2025).
41. Kodogiannis, V. S., Lygouras, J. N., Tarczynski, A. & Chowdrey, H. S. Artificial odor discrimination system using electronic nose and neural networks for the identification of urinary tract infection. *IEEE Trans. Inf Technol. Biomed.* **12** (6), 707–713. <https://doi.org/10.1109/TITB.2008.917928> (2008).
42. Komori, T., Nishikawa, H., Sasaki, K., Taniguchi, I. & Onoye, T. Multi-Class urinary sediment particles detection based on YOLOv7 with attention modules. *IEEE Access.* **12**, 129753–129764. <https://doi.org/10.1109/ACCESS.2024.3448262> (2024).
43. Lekha, S. & M, S. Recent advancements and future prospects on E-Nose sensors technology and machine learning approaches for Non-Invasive diabetes diagnosis: A review. *IEEE Rev. Biomed. Eng.* **14**, 127–138. <https://doi.org/10.1109/RBME.2020.2993591> (2021).
44. Massy, Z. A., Lambert, O., Metzger, M., Sedki, M., Chauvet, A., Breuil, B., Jaafar, A., Tack, I., Nguyen-Khoa, T., Alves, M., Siwy, J., Mischak, H., Verbeke, F., Glorieux, G., Herpe, Y.-E., Schanstra, J. P., Stengel, B., Klein, J., Alencar De Pinho, N., ... Speyer, E. (2023). Machine Learning-Based Urine Peptidome Analysis to Predict and Understand Mechanisms of Progression to Kidney Failure. *Kidney International Reports.* **8**(3), 544–555. <https://doi.org/10.1016/j.ekir.2022.11.023>.
45. McKay, G. N. et al. Lens free holographic imaging for urinary tract infection screening. *IEEE Trans. Biomed. Eng.* **70** (3), 1053–1061. <https://doi.org/10.1109/TBME.2022.3208220> (2023).
46. Mosterd, C. M., Verhaar, B. J. H., Van Den Born, B. J. H., Nieuwendorp, M. & Van Raalte, D. H. Plasma and urine metabolites associated with Non-diabetic chronic kidney disease: the HELIUS study. *Kidney Med.* 101009. <https://doi.org/10.1016/j.xkme.2025.101009> (2025).
47. Panda, A. R. et al. Kidney stone prediction based on urine analysis: A comprehensive study of machine learning models. *2025 Int. Conf. Emerg. Syst. Intell. Comput. (ESIC).* **748**–753 <https://doi.org/10.1109/ESIC64052.2025.1096264> (2025).
48. Rao, Z. et al. Construction of an ECL-DPV dual model biosensor for dopamine detection based on PSO-ANN algorithm. *IEEE Sens. J.* **24** (6), 7463–7472. <https://doi.org/10.1109/JSEN.2024.3358413> (2024).
49. Ritter, R. C., Zinner, N. R. & Sterling, A. M. The urinary drop Spectrometer—An electrooptical instrument for urological analysis based on the external urine stream. *IEEE Trans. Biomedical Eng. BME.* **-23** (3), 266–269. <https://doi.org/10.1109/TBME.1976.324642> (1976).
50. Roux-Dalvai, F. et al. Fast and accurate bacterial species identification in urine specimens using LC-MS/MS mass spectrometry and machine learning\*\*. *Mol. Cell. Proteom.* **18** (12), 2492–2505. <https://doi.org/10.1074/mcp.T1R119.001559> (2019).
51. S, S., K, L. & Technologies, N. A, R. K., A Comprehensive Analysis of Machine Learning Algorithms in Diagnosis of Chronic Kidney Disease. 2022 13th International Conference on Computing Communication and (ICCCNT), 1–5. (2022). <https://doi.org/10.1109/ICCCNT54827.2022.9984275>
52. Saberi, Z., Rezaei, B., Seyedhamzeh, M. & Ensafi, A. A. A novel aptasensor based on the formation of intermolecular G quadruplex DNA and carbon Dots for fluorescence determination potassium ions in human urine and blood serum samples. *IEEE Sens. J.* **21** (15), 16443–16450. <https://doi.org/10.1109/JSEN.2021.3077940> (2021).
53. Samet, S., Laouar, M. R., Bendib, I. & Systems, A. Use of Machine Learning Techniques to Predict Diabetes at an Early Stage. 2021 International Conference on Networking and (ICNAS), 1–6. (2021). <https://doi.org/10.1109/ICNAS53565.2021.9628903>
54. Seo, W., Yu, W., Tan, T., Ziae, B. & Jung, B. Diaper-Embedded urinary tract infection monitoring sensor module powered by Urine-Activated batteries. *IEEE Trans. Biomed. Circuits Syst.* **11** (3), 681–691. <https://doi.org/10.1109/TBCAS.2017.2654421> (2017).
55. Yadav, S. et al. A novel effective forecasting model developed using ensemble machine learning for early prognosis of asthma attack and risk grade analysis. *Scalable Computing: Pract. Experience.* **26** (1), 398–414 (2025).
56. Sheele, J. M., Campbell, R. L. & Jones, D. D. Machine learning to predict urine culture antibiotic sensitivities in the emergency department. *Heliyon* **11** (4), e42737. <https://doi.org/10.1016/j.heliyon.2025.e42737> (2025).
57. Shokhrukh, P. et al. Automated Kidney Disease Detection Using Machine Learning and Computer Vision Basedon Radiology Image Analysis. 2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS), 575–580. (2024). <https://doi.org/10.1109/ICTACS62700.2024.10840641>
58. Siegel, N. A. et al. Identifying patterns of tobacco use and associated cardiovascular disease risk through machine learning analysis of urine biomarkers. *JACC: Adv.* **4** (3), 101630. <https://doi.org/10.1016/j.jacadv.2025.101630> (2025).
59. Dalal, S. et al. Enhancing thyroid disease prediction with improved XGBoost model and bias management techniques. *Multimedia Tools Appl.* **84**, 1–32. (2024).
60. Smith, S., White, P., Redding, J., Ratcliffe, N. M. & Probert, C. S. J. Application of similarity coefficients to predict disease using volatile organic compounds. *IEEE Sens. J.* **10** (1), 92–96. <https://doi.org/10.1109/JSEN.2009.2035771> (2010).
61. Sonone, N. & Daniel, A. Leveraging Xgboost for Accurate Prediction of Chronic Kidney Disease with Real-Time Data. 2024 IEEE International Conference on Intelligent Signal Processing and Effective Communication Technologies (INSPECT), 1–6. (2024). <https://doi.org/10.1109/INSPECT63485.2024.10896080>
62. Swarup Kumar, J. N. V. R., Karri, C., Vatsavayi, N. S., Lekkala, H. & Choudarapu, S. Breast Cancer Detection Using Nanoparticle Sensor with Machine Learning Algorithms. 2023 4th International Conference on Intelligent Technologies (CONIT), 1–5. (2024). <https://doi.org/10.1109/CONIT61985.2024.10627465>
63. Thakur, R., Maheshwari, P., Datta, S. K., Dubey, S. K. & Shakher, C. Machine Learning-Based rapid Diagnostic-Test reader for albuminuria using smartphone. *IEEE Sens. J.* **21** (13), 14011–14026. <https://doi.org/10.1109/JSEN.2020.3034904> (2021).
64. Woodburn, E. V., Long, K. D. & Cunningham, B. T. Analysis of Paper-Based colorimetric assays with a smartphone spectrometer. *IEEE Sens. J.* **19** (2), 508–514. <https://doi.org/10.1109/JSEN.2018.2876631> (2019).
65. Ye, Z. T., Tseng, S. F. & Tsou, S. X. Detection and quantification of urine albumin using a miniature colorimetric sensor consisting of UVC leds and CsPbBr<sub>3</sub> quantum Dot color converters. *IEEE J. Sel. Top. Quantum Electron.* **30** (3: Flexible Optoelectronics), 1–7. <https://doi.org/10.1109/JSTQE.2024.3354387> (2024).
66. Yuting, Y. & Shan, D. Associations between urinary and blood heavy metal exposure and heart failure in elderly adults: insights from an interpretable machine learning model based on NHANES (2003–2020). *Int. J. Cardiol. Cardiovasc. Risk Prev.* **25**, 200418. <https://doi.org/10.1016/j.ijcrp.2025.200418> (2025).
67. Zhou, J., Welling, C. M., Vasquez, M. M., Grego, S. & Chakrabarty, K. Sensor-Array Optimization Based on Time-Series Data Analytics for Sanitation-Related Malodor Detection. *IEEE Trans. Biomed. Circuits Syst.* **14**(4), 705–714. <https://doi.org/10.1109/TBCAS.2020.3002180> (2020).
68. Carion, N. et al. End-to-End object detection with Transformers. In *Computer Vision – ECCV 2020. ECCV 2020. Lecture Notes in Computer Science() Vol. 12346* (eds Vedaldi, A., Bischof, H., Brox, T., Frahm, J. M. et al.) (Springer, 2020). [https://doi.org/10.1007/978-3-030-58452-8\\_13](https://doi.org/10.1007/978-3-030-58452-8_13)
69. Sharma, P. & Sharani, P. (2014). Design of photonic crystal-based biosensor for detection of glucose concentration in urine. *IEEE Sensors Journal*, **15**(2), 1035–1042.
70. Al Sayed, T. O., Essam, M. D., & Al Saeed, K. O. (2024). Evaluation of Tamsulosin 0.4 mg versus 0.8 mg in Management of Patients Suffering from Lower Urinary Tract Symptoms Due to Benign Prostatic Hyperplasia. *QJM: An International Journal of Medicine*, **117**(Supplement\_1), hcae070–628.

71. Deng, C., Zhou, X., Zhang, L., You, Q., Liu, C., Zhang, Y., & Yang, J. (2025). Adipokine isthmin-1 is a potential predictor of abnormal urine Na<sup>+</sup> excretion and insulin resistance for primary hypertension. *BMC Cardiovascular Disorders*, **25**(1), 136.
72. Durán Acevedo, C. M., Carrillo Gómez, J. K., Cuastumal Vasquez, C. A., & Ramos, J. (2024). Prostate cancer detection in Colombian patients through E-senses devices in exhaled breath and urine samples. *Chemosensors*, **12**(1), 11.

## Acknowledgments

This research was funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2025R761), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

## Author contributions

ND, DP & SK were responsible for Validation, Software, Data Curation, and Writing - Original Draft. SRK & IS were responsible for Conceptualization, Writing - Original Draft. MAS & MD were responsible for Writing the original Draft and visualization. MD was responsible for Writing - Review & Editing. MA & AH were responsible for Formal Analysis. MA, SoK & MD were responsible for Writing - Original Draft, Resources, Supervision. The author(s) read and approved the final manuscript.

## Funding

This research has received no funding.

## Declarations

### Competing interests

The authors declare no competing interests.

### Conflict of interest

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

### Additional information

Correspondence and requests for materials should be addressed to M.D.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025