scientific reports

Check for updates

OPEN

A multi-head YOLOv12 with self-supervised pretraining for urinary sediment particle detection

Mehdi Alizadeh[1], Ali Karimi[2], Mohammad Javad Barikbin[2], Ali Movahed[3], Samad Akbarzadeh[3], Majid Sirati-Sabet[1]✉ & Mohammad Ali Akhaee[2]✉

Automatic and reliable urine sediment analysis is essential for timely diagnosis and management of renal and urinary disorders. However, manual methods are time-consuming, subjective, and limited by operator abilities. In this study, we propose a novel deep learning method based on a multi-head YOLOv12 architecture combined with self-supervised pretraining and advanced inference through Slicing Aided Hyper Inference (SAHI) to effectively address these challenges. Unlike prior methods that employed a single detection head, our architecture features six specialized and independent detection heads: Cells, Casts, Crystals, Microorganisms/Yeast, Artifact, and Others, enabling simultaneous and fine-grained classification of the full spectrum of urine sediment particles, including all relevant subclasses. To facilitate robust training, we created a large-scale dataset (OpenUrine) encompassing 790 labeled images with over 31,285 bounding boxes across 39 categories, and 5640 unlabeled images for self-supervised learning. Evaluated on this complex 39-class dataset, our model achieved a precision of 76.59% and a mean Average Precision (mAP) of 64.15%, demonstrating competitive performance in detection accuracy, especially of small and low-contrast objects.

Urine sediment analysis is a cornerstone of clinical diagnostics, essential for the assessment and management of kidney diseases, urinary tract infections, and various systemic disorders[1,2]. In clinical practice, examining the microscopic components of urine, such as cells, casts, and crystals, provides vital clues about renal function and the presence of pathological abnormalities[1]. However, despite its clinical significance, manual urine sediment analysis remains a labor-intensive, subjective, and operator-dependent procedure, which leads to variability in results between professionals and laboratories[3–5]. With the rising volume of laboratory requests and limited availability of skilled personnel, there is a pressing need for accurate and efficient automated methods that deliver reliable and standardized results[4].

The main problem addressed in our study is the challenge of automating urine sediment analysis using artificial intelligence (AI), especially for detecting and classifying the full spectrum of urinary particles. Many of these particles have highly diverse morphologies, small sizes, and are frequently underrepresented in datasets[6,7]. Although state-of-the-art AI methods are promising, they often depend on large, labeled datasets and tend to struggle with real-world image diversity and rare category detection[5,6,8]. This situation highlights the need for novel solutions that can leverage both labeled and unlabeled data within robust architectures.

Deep learning, especially convolutional neural networks (CNNs), has revolutionized medical image analysis in recent years. It offers automated feature extraction and remarkable performance in tasks such as disease detection, localization, and segmentation across multiple imaging modalities[9–12]. However, applying deep learning to urine sediment images presents unique challenges, including the lack of large annotated datasets, high-resolution imaging requirements, and wide variation in image quality. Recent advances demonstrate the value of self-supervised learning[13,14], where unlabeled data is used to pretrain models via methods such as image reconstruction. This method can help to overcome data scarcity and improve model generalizability[15]. In this context, we introduce a large and diverse dataset, OpenUrine, which contains 790 labeled images (with over 31,285 expert-annotated bounding boxes across 39 categories) and an additional 5,640 unlabeled images for self-supervised learning.

A key innovation in our study is the design of a multi-head YOLOv12 architecture. Six parallel detection heads are specifically dedicated to Cells, Casts, Crystals, Microorganisms/Yeast, Artifacts, and Others. These heads operate simultaneously to enable comprehensive and precise detection of all relevant urinary sediment

[1]Department of Clinical Biochemistry, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran. [2]School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran. [3]Department of Clinical Biochemistry, Bushehr University of Medical Sciences, Bushehr, Iran. ✉email: sirati@sbmu.ac.ir; akhaee@ut.ac.ir

particles and their respective subclasses. Unlike previous single-head models, this architecture allows the model to independently capture the distinct morphological and visual characteristics of diverse particle types. Such a multi-head mechanism is essential for robust identification and discrimination among particle classes that differ widely in size, shape, and appearance, ensuring that both common and rare elements are detected with high accuracy.

The primary goal of this research is to develop and validate an effective and scalable deep learning method based on the multi-head YOLOv12, complemented by self-supervised pretraining and Slicing Aided Hyper Inference (SAHI)-based inference, for comprehensive and automated detection and classification of urinary sediment particles in microscopy images.

Related works

The potential of deep learning to overcome the limitations of traditional urine sediment analysis has spurred significant research efforts in developing deep learning-based AI models for automated analysis[6]. These models are designed to automatically identify and classify the various microscopic particles found in urine sediment, including red blood cells, white blood cells, epithelial cells, casts, crystals, bacteria, and yeast. Researchers have explored a wide range of deep learning architectures for this purpose, with CNNs being particularly prominent. Models such as AlexNet[16], ResNet[17], GoogleNet[18], DenseNet[19], MobileNet[20], and YOLO have been adapted and applied to the task of urine particle classification and detection.

Some studies have focused on specific clinical applications of these AI models. For instance, research has explored the use of deep learning to detect bacteria in urine samples directly from microscopic images, eliminating the need for traditional, time-consuming urine culture methods[21]. Another study has investigated the potential of AI to screen for rare diseases, such as Fabry disease, by identifying unique cellular morphologies in urine sediment images[22]. To enhance the performance of these models, researchers are continuously exploring various techniques, including novel image amplification methods to augment training datasets, the incorporation of attention mechanisms to focus on relevant image features, and the development of hybrid methods that combine the strengths of CNNs with traditional feature extraction techniques like Local Binary Patterns (LBP)[5,6,23,24]. The use of pre-trained models and transfer learning is also a common strategy, allowing researchers to leverage knowledge gained from training on large general image datasets to improve the performance of models on the often-smaller urine sediment image datasets[8] Furthermore, object detection methods like Faster R-CNN[25], SSD[26], and YOLO are being applied to simultaneously locate and classify urine particles within microscopic images, providing a more comprehensive analysis than simple image-level classification[6].

The application of deep learning to urine sediment analysis encompasses various methodes tailored to specific analytical needs. Many studies focus on classification tasks, where the goal is to categorize individual urine sediment particles into predefined classes, such as red blood cells, white blood cells, and different types of crystals[5,12]. These models learn to recognize the distinct visual features of each particle type to perform accurate classification. Another significant method involves object detection tasks, where the AI model aims to not only classify but also to precisely locate multiple urine particles within a single microscopic image[7,12]. This is particularly valuable in clinical settings as it allows for the quantification of different particle types and the analysis of their spatial relationships within the urine sediment.

Liang et al.[27] in their study used a dataset containing 10,752 images with seven classes consisting of urinary particles (erythrocytes, leukocytes, epithelial, low-transitional epithelium, casts, crystal, and squamous epithelial cells). It was stated that after balancing the image categories, the data was used to train a RetinaNet model[28]. It was stated that an 88.65% accuracy value was obtained with this developed method on a test set, with a processing time of 0.2 s per image. Yildirim et al.[5] in their study used a data set containing 8,509 particle images with eight classes obtained from urine sediment. They developed a hybrid model based on textural (LBP) and ResNet50. It was stated that after optimizing and combining features, a high accuracy value of 96.0% was obtained with the proposed model. Liang et al.[23] conducted a series of studies aimed at improving urinary sediment analysis through deep learning-based object detection models. In one study, they proposed the Dense Feature Pyramid Network (DPFN) architecture, integrating DenseNet into the standard FPN model and incorporating attention mechanisms into the network head. This method significantly mitigated class confusion in urine sediment images, particularly improving erythrocyte detection accuracy from 65.4% to 93.8%, and achieving a mean average precision (mAP) of 86.9% on the test set. In a complementary study[29], they framed urinary particle recognition as an object detection task using CNN-based models such as Faster R-CNN and SSD. Evaluated on a dataset of 5,376 labeled images across seven urinary particle categories, their best-performing model achieved an mAP of 84.1%.

Ji et al.[15] proposed a semi-supervised network model (US-RepNet) to classify urine sediment images. They used a data set containing 429,605 urine sediment images with 16 classes. They stated that they obtained a 94% accuracy value with their suggested model. Li et al.[30] in their study used a data set containing 2551 urine sediment images with four classes (red blood cells, white blood cells, epithelial cells, and crystals). They developed a modified LeNet-5[31]. They stated that they performed classification with 92% accuracy. Khalid et al.[24] compiled a dataset of 820 annotated urine sediment images. This dataset was used to train and evaluate five convolutional neural network models - MobileNet, VGG16[32], DenseNet, ResNet50, and InceptionV3[33] - along with a proposed CNN architecture. MobileNet achieved the highest true positive recall, followed closely by the proposed model. Both models reached a top accuracy of 98.3%, while InceptionV3 and DenseNet demonstrated slightly lower but still comparable accuracy of 96.5

Avci et al.[34] developed a model for urinary particle recognition that enhances the resolution of microscopic images using a super-resolution Faster R-CNN method. They utilized pre-trained architectures including AlexNet, VGG16, and VGG19. Among these, the AlexNet-based model delivered the best performance, achieving a recognition accuracy of 98.6%. In another study[35], they introduced a combination of Discrete

Wavelet Transform (DWT) and a neural network-based system, the ADWEENN algorithm, for recognizing 10 different categories of urine sediment particles, achieving an accuracy of 97.58%.

In another study, Erten et al. introduced Swin-LBP, a handcrafted feature engineering model for urine sediment classification that combines the Swin transformer architecture with local binary pattern (LBP) techniques. Their six-phase approach—including LBP-based feature extraction, neighborhood component analysis (NCA) for feature selection, and support vector machine (SVM)[36] classification achieved an accuracy of 92.60% across 7 classes of urinary sediment elements, outperforming conventional deep learning methods applied on the same dataset[37]. In a subsequent study, the same group proposed another model integrating cryptographic-inspired image preprocessing techniques, notably the Arnold Cat Map (ACM), with patch-based mixing and transfer learning. Leveraging DenseNet201 for deep feature extraction and NCA for feature selection, this model reached an even higher classification accuracy of 98.52% for seven types of urinary particles[8].

A recent study proposed a combined CNN model integrated with an Area Feature Algorithm (AFA), enabling improved recognition of 10 urine sediment categories from a large dataset of 300,000 images, achieving a test accuracy of 97% and significantly enhancing the recognition of visually similar particles such as RBCs and WBCs[38]. A deep learning model based on VGG-16 was developed to classify 15 types of urinary sediment crystals using 441 images, which were augmented to 60,000 images through targeted data augmentation. Removing the random cropping step in data augmentation significantly improved accuracy, and the model achieved a performance of 91.8%[39].

Lyu et al.[7] developed an advanced deep learning model, YUS-Net, based on an improved YOLOX[40] architecture for multi-class detection of urinary sediment particles. The model integrates domain-specific data augmentation, attention mechanisms, and Varifocal loss to enhance the detection of challenging particle types, particularly small and densely distributed objects. Evaluated on the USE dataset, YUS-Net achieved impressive performance, with a mean Average Precision (mAP) of 96.07%, 99.35% average precision, and 96.77% average recall, demonstrating its potential for efficient and accurate end-to-end urine sediment analysis.

A critical limitation of existing research is the narrow scope of detection. The vast majority of published object detection studies focus on a small number of classes. For example, the influential work by Liang et al.[29] used a dataset of 5,376 labeled images across seven urinary particle categories. The dataset used by Li et al.[27] also contained seven classes. The hybrid classification model by Yildirim et al.[5] was trained on eight particle types. Even more ambitious studies, such as that by Ji et al.[15], which used a large dataset, topped out at 16 categories.

Beyond prior urine microscopy studies, several recent deep learning frameworks across other domains further highlight the rapid evolution of hybrid architectures. In biomedical imaging, models such as DCSSGA-UNet[41] and EFFResNet-ViT[42] adopt dense connectivity, semantic attention, and CNN–Transformer fusion to enhance segmentation and classification precision. Similarly, deep hybrid and self-supervised architectures from cyber-physical security research[43–45] demonstrate parallel methodological advances in representation learning and encoder–decoder design. Comparable trends have also appeared in unrelated areas such as sports performance analytics and wearable sensor forecasting[46,47], reflecting the general shift toward multi-branch and attention-driven deep models across domains.

This "granularity gap" between existing research and the diverse reality of clinical samples is a major barrier to practical deployment. Our work directly confronts this gap by introducing a model and a public dataset, OpenUrine, designed for the comprehensive detection of 39 distinct categories, representing a significant leap in complexity and clinical relevance.

Dataset

The dataset utilized in this study, named OpenUrine, comprises 6430 images of the urinary sediment. OpenUrine consists of a total of 790 anonymized, expert-labeled microscopic images of urinary sediment, in addition to 5,640 unlabeled images used for self-supervised learning. This is the first publicly available dataset dedicated to urinary particle detection. No patient metadata was collected at any stage; all samples were fully anonymized and are referenced only by randomly assigned identification codes. None of the images carry patient-specific information, ensuring complete privacy and compliance with ethical data standards. Images were collected from multiple laboratories using different microscope models and various smartphone cameras to ensure a broad range of imaging conditions reflective of real-world clinical variability.

An overview of the dataset, including the number of labeled and unlabeled images as well as the total number of bounding box annotations, is presented in Table 1. Table 2 provides a detailed breakdown of all 39 categories, reporting the number of annotated objects, number of images containing each label, and a brief scientific description for each particle type, facilitating a comprehensive understanding of the dataset's diversity and clinical relevance.

Dataset	Image	Box
Labeled data	790	31,285
Unlabeled data	5640	–

Table 1. Summary of the OpenUrine dataset, detailing the number of images and annotated bounding boxes for both labeled and unlabeled subsets.

Label	Category	Number of boxes	Number of images	Description
Muddy brown cast	Casts	350	24	Brown casts indicating acute tubular necrosis; typically associated with kidney injury
Granular cast	Casts	172	63	Cylindrical structures formed from protein and cellular debris, indicating kidney damage
Hyaline cast	Casts	160	42	Clear, colorless casts formed from protein; may indicate dehydration or kidney disease
Waxy cast	Casts	56	42	Broad casts indicating chronic renal failure; they appear broad and are formed from degenerated cells
Mixed cell cast	Casts	44	41	Casts containing various types of cells; their presence can indicate kidney pathology
RBC cast	Casts	20	19	Casts formed from red blood cells; presence indicates glomerular damage or bleeding within the kidneys
RBC	Cells	5752	213	Red blood cells; their presence in urine (hematuria) can indicate bleeding within the urinary tract
WBC	Cells	2654	210	White blood cells; an increased number (pyuria) suggests infection or inflammation, such as a urinary tract infection
Epithelial cell (squamous)	Cells	889	175	Flat cells from the urethra or vaginal lining; usually considered normal unless present in excess
Epithelial cell (transitional)	Cells	233	70	Cells lining the bladder and ureters; increased numbers may suggest infection, inflammation, or, rarely, malignancy
Suspected atypical cell	Cells	187	44	Unusual cells that may require further investigation to rule out malignancy.
Epithelial cell (renal)	Cells	126	46	Cells originating from the renal tubules; their presence may indicate tubular injury or necrosis
WBC clump	Cells	103	16	Aggregates of white blood cells; may indicate significant infection or inflammation within the urinary tract
Lipid cast	Cells	24	22	Casts containing fat droplets, indicating nephrotic syndrome when present in excess
RBC clump	Cells	24	5	Aggregates of red blood cells; may indicate significant bleeding or injury within the urinary tract
Calcium oxalate	Crystals	2872	115	Colorless, envelope-shaped (dihydrate) or dumbbell-shaped (monohydrate) crystals; commonly seen in individuals who consume oxalate-rich foods or have kidney stones
Amorphous	Crystals	951	35	Aggregates of fine granules; amorphous urates appear in acidic urine, while amorphous phosphates appear in alkaline urine; generally of little clinical significance
Triple phosphate	Crystals	778	31	Coffin-lid shaped crystals; typically found in alkaline urine and associated with urinary tract infections
Uric acid	Crystals	699	42	Yellow to reddish-brown, diamond or barrel-shaped crystals; common in acidic urine and can be associated with gout or chemotherapy
Ammonium biurat	Crystals	394	27	Yellow-brown, thorn-apple shaped crystals; often found in old or poorly preserved urine samples
Hippuric acid	Crystals	210	8	Colorless or pale yellow needles or prisms; rare and typically of little clinical significance
Calcium phosphate	Crystals	200	27	Colorless, needle-like or rosette formations; typically found in alkaline urine; generally not clinically significant
Cystine	Crystals	137	11	Colorless, hexagonal plates; indicative of cystinuria, a rare genetic disorder
Leucine	Crystals	91	11	Spherical crystals with concentric rings and radial striations; associated with severe liver disease
Tyrosine crystal	Crystals	81	20	Fine, needle-like crystals; may indicate severe liver disease
Calcium carbonate	Crystals	76	4	Small, colorless granules or dumbbell-shaped crystals; typically found in alkaline urine; usually not clinically significant
Cholesterol	Crystals	58	13	Large, flat, transparent plates with notched corners; may be seen in nephrotic syndrome.
Bilirubin	Crystals	43	11	Yellow to reddish brown, needle-like or granular crystals, associated with liver disorders
Bacteria	Microorganisms	12406	248	Their presence, especially in large numbers, suggests a urinary tract infection
Yeast	Microorganisms	446	32	Often appear as budding cells; can indicate a yeast infection, particularly in diabetic patients
Enterobius vermicularis egg	Microorganisms	81	13	Oval-shaped eggs with a characteristic flattened side; indicates pinworm infection, often due to fecal contamination
Fungal hyphae	Microorganisms	72	10	Branching filamentous structures; indicate a fungal infection, more common in immunocompromised individuals
Schistosoma haematobium eggs	Microorganisms	31	17	Oval eggs with a terminal spine; indicate schistosomiasis, a parasitic infection affecting the urinary tract
Trichomonas vaginalis	Microorganisms	26	17	A motile parasite; its presence indicates trichomoniasis, a sexually transmitted infection
Sperm	Others	382	21	May be present in urine after ejaculation; typically not clinically significant
Mucus	Others	198	59	Thread-like structures; generally not clinically significant but can be confused with casts
Fat droplets	Others	53	15	Free-floating droplets; indicate lipiduria, often associated with nephrotic syndrome
Oval fat bodies	Others	38	20	Renal tubular cells filled with fat droplets; suggestive of nephrotic syndrome
Artifact	Artifact	168	94	Any foreign substance or error that does not represent actual urine components
Total		31285	790	

Table 2. Detailed breakdown of the 39 categories in OpenUrine dataset. The total count represents image-label pairs, as individual images may contain multiple particle types. The actual number of unique images is 790.

Data labeling

Each image was assigned a unique identification code upon acquisition. Two experienced clinical biochemistry experts conducted the labeling process independently, ensuring high reliability and consensus in recognizing and delineating all urinary sediment structures present. All detectable objects were marked with bounding boxes and assigned one of the 39 class labels. Figure 1 presents sample annotated microscopic fields from the OpenUrine

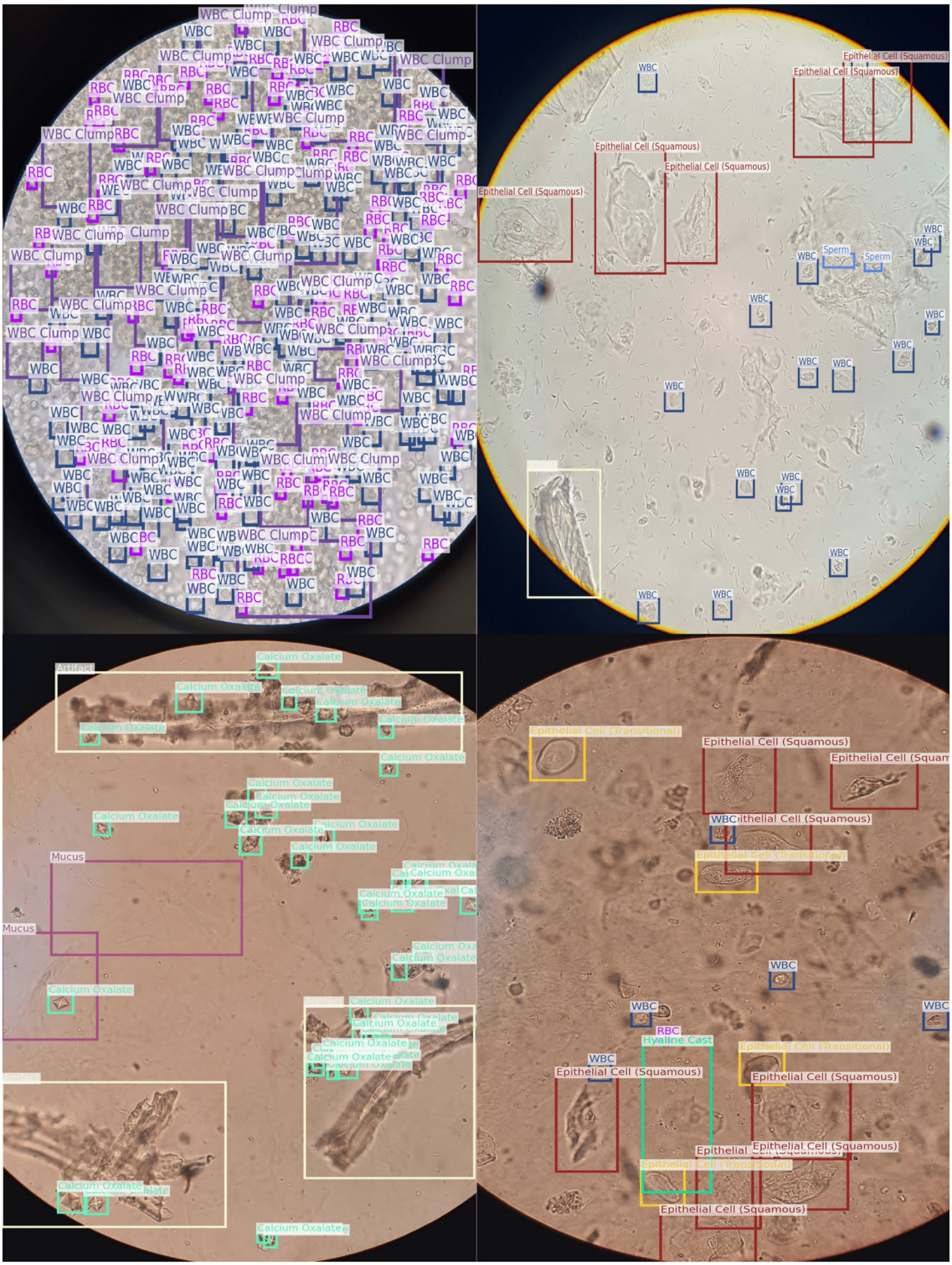

Fig. 1. Representative annotated microscopic fields from the OpenUrine dataset. Each sub-image shows a clinical urine sample with expert-labeled bounding boxes over multiple particle types, reflecting the high density, diversity, and spatial complexity encountered in real-world urinalysis.

dataset. Each sub-image shows a real clinical sample with expert-verified bounding box annotations identifying and localizing multiple urinary particles across diverse imaging conditions. Figure 2 displays representative examples of all 39 particle categories present in the dataset. Each image illustrates the unique morphology and appearance of a specific urinary sediment particle, such as various cell types, casts, crystals, microorganisms, and artifacts.

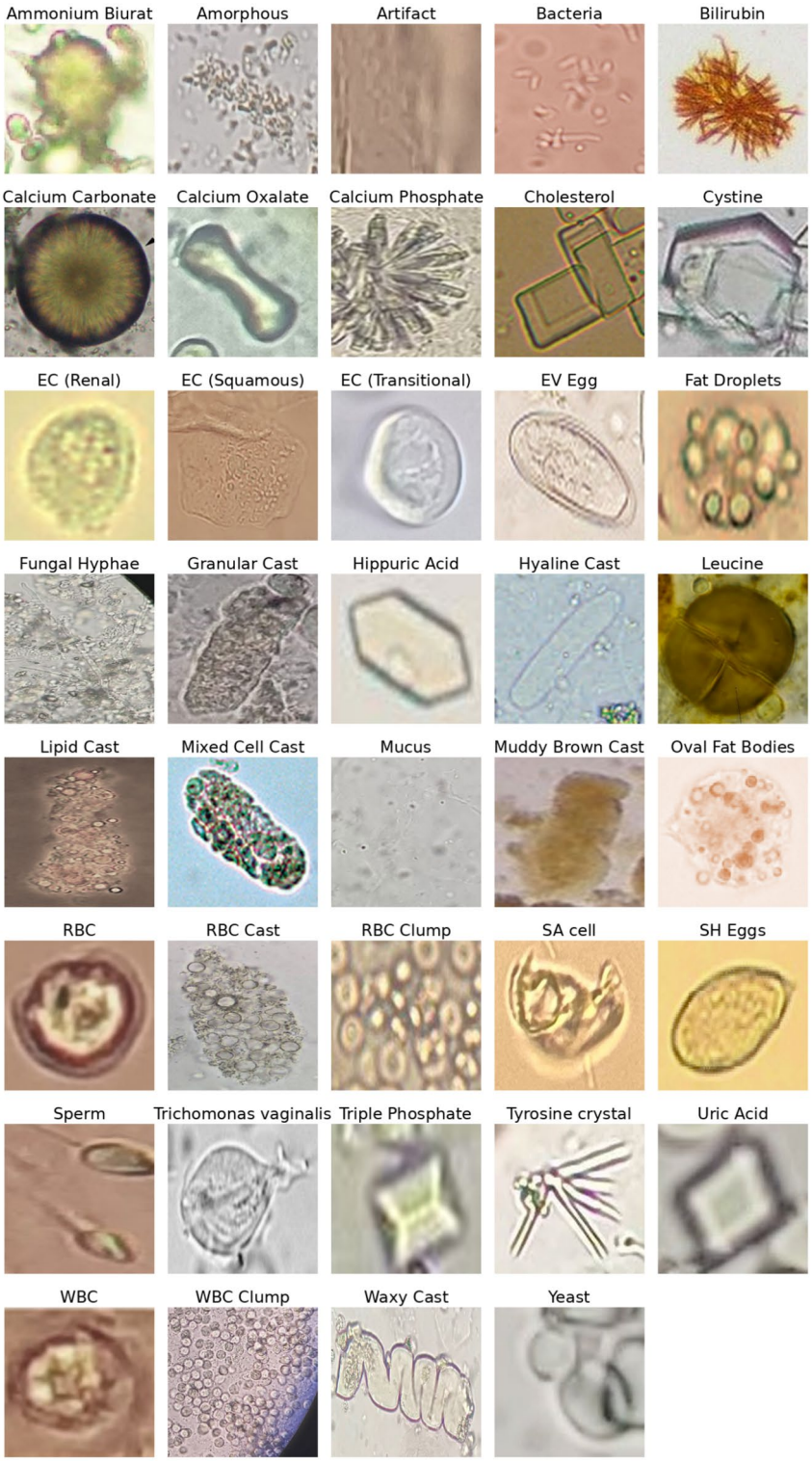

Fig. 2. Example images representing all 39 urinary sediment particle categories included in the OpenUrine dataset.

Unlabeled images for self-supervised learning

Beyond the labeled portion, the OpenUrine dataset also includes 5,640 unlabeled images. These images, which share the same acquisition characteristics as the labeled set, were used in the self-supervised stage of the proposed method to further boost model performance and robustness.

Data partitioning

The labeled dataset was divided into training and testing sets in an 80:20 ratio at the patient level, ensuring that all images from a single patient are assigned to either the training or testing set, but not both. This patient-level split prevents data leakage and ensures realistic evaluation of the model's generalization capability to new patients.

The labeled dataset was divided using 5-fold cross-validation. Each fold was trained independently, and the reported results represent the mean±std (%) across the five folds. This partitioning and validation process ensures fair and objective model assessment.

Method

This section outlines the methodology for fully automated detection and categorization of urinary sediment particles in high-resolution microscopy, leveraging a custom multi-head YOLOv12 architecture designed specifically for the OpenUrine dataset.

Architecture overview

As illustrated in Fig. 3, the proposed method is a multi-head object detection based on YOLOv12[48,49], adapted and optimized for challenging urinary sediment images (average size 1800×1800 px). A key innovation is the separation of the detection module into six distinct semantic heads, each corresponding to a clinically relevant super-category of urinary sediment objects. This structure enhances discrimination and robustness, particularly for rare or visually subtle subclasses. To further address the challenges of detecting small, densely packed structures in large fields, Slicing Aided Hyper-Inference (SAHI)[50] is tightly integrated into the inference pipeline.

Backbone network

Each input image X is processed by a YOLO backbone, which extracts multiscale, high-level feature maps:

$$F = \text{Backbone}(X)$$

These feature maps provide rich spatial and morphological representations crucial for accurate detection across a wide range of object scales.

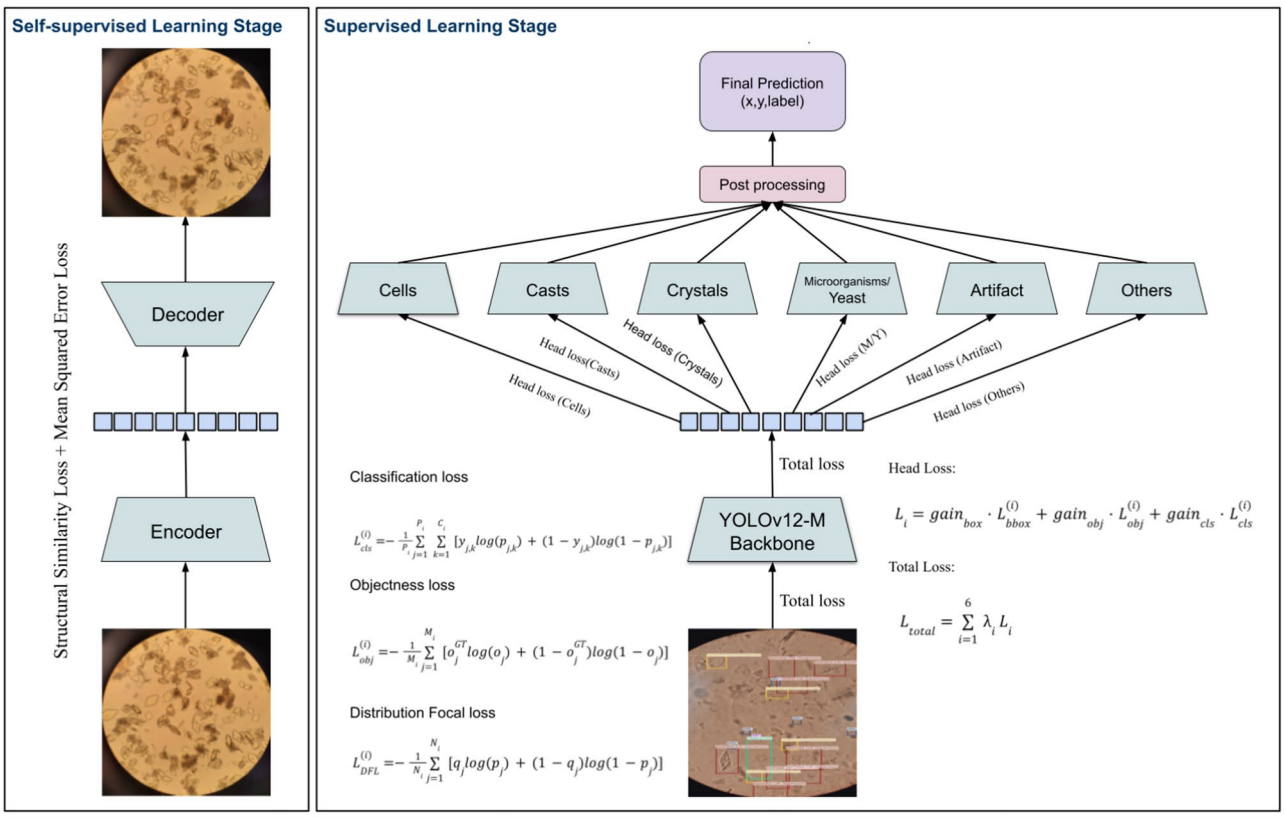

Fig. 3. Overview of the proposed two-stage deep learning method. (**A**) The encoder–decoder network is pretrained via self-supervised reconstruction on unlabeled urine sediment images to learn rich feature representations. (**B**) The pretrained encoder (backbone) is fine-tuned for object detection using six parallel heads, enabling precise multi-class identification of urinary particles.

Multi-head detection module

The detection module utilizes six parallel output heads, each specializing in one clinically important super-category of urinary sediment particles: Cells, Casts, Crystals, Microorganisms/Yeast, Artifact, and Others. This categorization directly follows established clinical taxonomy and precisely matches the semantic groupings defined in Table 2. Each head is responsible for detecting all subcategories corresponding to its group.

For each group i, the shared feature map F is passed to the corresponding detection head:

$$Y_i = \mathrm{Head}_i(F)$$

where Y_i encodes the bounding boxes, objectness scores, and class probabilities for all subclasses assigned to that head.

Loss function

In our architecture, the detection loss follows the YOLOv12 formulation[48], but is applied independently to each of the six output heads. This design allows every head, specialized for its own clinical super-category, to optimize its parameters without interference from unrelated particle types, while still contributing to the overall network performance.

The total loss is the weighted sum of the head-specific losses, as shown in Equation 1.

$$L_{\text{total}} = \sum_{i=1}^{6} \lambda_i \, L_i \tag{1}$$

where λ_i controls the relative weight of head i based on its clinical importance and representation in the dataset. The values are tuned as shown in Table 3.

Each head-specific loss L_i is defined in Eq. 2.

$$L_i = \text{gain}_{\text{box}} \cdot L_{\text{bbox}}^{(i)} + \text{gain}_{\text{obj}} \cdot L_{\text{obj}}^{(i)} + \text{gain}_{\text{cls}} \cdot L_{\text{cls}}^{(i)} \tag{2}$$

where gain_{box}, gain_{cls}, and gain_{obj} correspond to box loss gain, classification loss gain, and objectness scaling hyperparameters defined in Table 3.

Bounding box regression in YOLOv12 combines IoU-based loss[51] with the Distribution Focal Loss (DFL)[52] to enhance localization precision. The bounding box regression loss for head i is expressed in Eq. (3).

$$L_{\text{bbox}}^{(i)} = L_{\text{CIoU}}^{(i)} + \text{gain}_{\text{DFL}} \cdot L_{\text{DFL}}^{(i)} \tag{3}$$

Here, L_{CIoU} accounts for overlap, center distance, and aspect ratio, while L_{DFL} refines predicted box coordinates at sub-pixel resolution. The complete IoU loss term is defined in Eq. (4).

Parameter	Optimal value
Input image size	(960,960)
Optimizer	SGD with momentum
Momentum	0.939
Initial learning rate (lr0)	0.00996
Final learning rate (lrf)	0.00724
Loss weight coefficients (λ)	Artifact=0.5, casts=1.5, others=1.5, other heads=1.0
Weight decay	0.0005
Warmup epochs	2.03
Warmup momentum	0.874
Box loss gain	5.5
Classification loss gain (cls)	0.699
DFL loss gain	1.43
HSV_h (hue)	0.0173
HSV_s (saturation)	0.828
HSV_v (value)	0.238
Translate (fraction)	0.135
Scale (gain factor)	0.35
Fliplr (horizontal flip)	0.286
Mosaic	0.408

Table 3. Optimized hyperparameters for our method training on the OpenUrine dataset.

$$L_{\text{CIoU}}^{(i)} = 1 - \text{IoU}(b, b^{GT}) + \frac{\rho^2(b_c, b_c^{GT})}{c^2} + \alpha v \tag{4}$$

where ρ is the center-point distance, c is the diagonal length of the smallest enclosing box, and v is the aspect ratio term with balance factor α. The distribution focal loss is expressed in Eq. (5).

$$L_{\text{DFL}}^{(i)} = -\frac{1}{N_i} \sum_{j=1}^{N_i} \left[q_j \log(p_j) + (1 - q_j) \log(1 - p_j) \right] \tag{5}$$

where p_j is the predicted probability for the discretized bin of a coordinate value and q_j is the corresponding soft target.

Objectness loss[53] measures how well the model distinguishes objects from background. The objectness loss for head *i* is formulated in Eq. (6).

$$L_{\text{obj}}^{(i)} = -\frac{1}{M_i} \sum_{j=1}^{M_i} \left[o_j^{GT} \log(o_j) + (1 - o_j^{GT}) \log(1 - o_j) \right] \tag{6}$$

where o_j is the predicted objectness score for anchor *j*, and $o_j^{GT} \in \{0, 1\}$.

Classification loss[53] ensures correct subclass identification within each head. The classification loss for head *i* is defined in Eq. (7).

$$L_{\text{cls}}^{(i)} = -\frac{1}{P_i} \sum_{j=1}^{P_i} \sum_{k=1}^{C_i} \left[y_{j,k} \log(p_{j,k}) + (1 - y_{j,k}) \log(1 - p_{j,k}) \right] \tag{7}$$

where $p_{j,k}$ is the predicted probability for subclass *k* in sample *j*, and C_i is the number of subclasses in head *i*.

Unlike a unified detector that learns all categories together, here each head focuses only on the visual patterns of its assigned group, using Eqs. 3 through 7 independently. This separation avoids competition between unrelated classes, reduces the impact of severe class imbalance, and allows adjusting λ_i in Eq. 1 to boost underrepresented yet clinically significant categories. As our ablation studies show, removing this head-level independence leads to the sharpest drop in mean Average Precision (mAP) and recall.

Training procedure

A two-stage training strategy, optimized to leverage both labeled and unlabeled data, is applied.

(1) Self-supervised pretraining: during the self-supervised pretraining stage, all 5640 unlabeled images were utilized in an image reconstruction autoencoder architecture (as illustrated in Fig. 3). The network follows an encoder-decoder structure: the encoder mirrors the YOLOv12 backbone to extract latent morphological representations, and the decoder reconstructs the input image using these features. The model was optimized with a combined L1 + SSIM reconstruction loss, enforcing both pixel-level accuracy and structural consistency between input and reconstructed outputs. This pretext task effectively encourages the backbone to capture intrinsic microscopic texture and morphology priors even without labels. The pretrained encoder weights were subsequently transferred to initialize the YOLOv12 backbone during the supervised fine-tuning stage.

(2) Supervised fine-tuning: the backbone's pretrained weights initialize the detection model, which is then fine-tuned using the 790 image-level-labeled samples and bounding box annotations. Each particle is routed to its corresponding semantic head, and the total loss is jointly optimized. Diverse data augmentation (e.g., Mosaic, Sacle) and a SGD with Momentum are employed.

Inference with SAHI

For inference, we employed SAHI to enhance detection performance on high-resolution microscopic images. SAHI systematically divides each input image into overlapping tiles of 640×640 pixels with an overlap ratio of 0.25 (25%) in both horizontal and vertical directions. This slicing strategy enables the model to process smaller image regions with higher effective resolution, significantly improving detection sensitivity for small and densely packed urinary particles that might be missed in full-resolution inference.

Each slice is independently processed by our proposed method, generating separate predictions for particles within that region. The tiled outputs are subsequently merged through non-maximum suppression (NMS) to eliminate duplicate detections and produce consolidated, non-redundant bounding boxes.

Results

To comprehensively evaluate the effectiveness of our proposed object detection method, we conducted a series of experiments on the OpenUrine dataset. These experiments were specifically designed to demonstrate the superiority of our method compared to prior object detection methods under identical conditions.

Performance evaluation metrics

Model performance was assessed using several established object detection metrics. Precision quantifies the proportion of correctly identified positive detections among all predicted positives, as defined in Eq. (8):

$$\text{Precision} = \frac{TP}{TP + FP} \tag{8}$$

where TP and FP denote the numbers of true positive and false positive predictions, respectively. Recall measures the proportion of actual positives that are correctly detected by the model, as shown in Eq. (9):

$$\text{Recall} = \frac{TP}{TP + FN} \tag{9}$$

where FN is the number of false negatives. The overall detection capability is further summarized by the mean Average Precision (mAP), which is the unweighted mean of the Average Precision (AP) across all object classes, as presented in Eq. (10):

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^{N} \text{AP}_i \tag{10}$$

where N is the total number of classes under consideration. Additionally, the evaluation follows the COCO protocol[54] by reporting mAP@50-95, which represents the mean AP computed over multiple intersection-over-union (IoU) thresholds ranging from 0.5 to 0.95 (in increments of 0.05), thereby providing a stricter and more comprehensive measure of detection performance.

Implementation details

A comprehensive hyperparameter optimization protocol was carried out as part of our experimental design (see Fig. 4). For this purpose, we performed 300 independent training runs, each for 100 epochs, gradually searching the space of learning rate, momentum, weight decay, and various augmentation factors as listed in Table 3. Each configuration was evaluated on the validation split after every epoch, allowing us to systematically identify optimal values. All experiments, including baseline comparisons, were performed on the OpenUrine dataset for scientific consistency.

The scatter plots in Fig. 4 visualize the relationship between key hyperparameters and resulting detection metrics (such as mAP, mAP@50-95, Precision, and Recall); final selected values are denoted by a cross marker.

For the final training of our best-performing model, we utilized the optimal parameters over 300 epochs with a batch size of 16 and an input image size of 960×960 pixels, ensuring maximal capacity to learn robust object representations.

Comparative evaluation

Quantitative and qualitative evaluation of automated urine sediment analysis models is crucial for establishing their accuracy, robustness, and clinical viability. In this section, we present a comprehensive comparative analysis of our proposed method,and state-of-the-art methods, followed by investigations into how input image size and particle class influence model performance. The reliability and interpretability of the deep network are further validated through visual explanation techniques such as Grad-CAM.

Comparison with state-of-the-art methods

Table 4 provides a comprehensive comparison between our method, ans state-of-the-art methods. Our proposed method achieves the highest performance on all core metrics (precision, recall, mAP_{50}, mAP_{50-95}), outperforming both the latest YOLO models and prior state-of-the-art methods. While absolute values such as 76.59% precision may appear modest compared to simpler tasks, it is important to note that the OpenUrine dataset includes 39 diverse classes, making it a far more complex challenge than datasets used in previous studies. The ablation results reveal that both the multi-head detection strategy and self-supervised pretraining contribute substantially to the observed gains. In particular, removing the multi-head scheme leads to the largest drop in precision, recall, and mean average precision, highlighting the value of specialized detection branches for different particle types. Our method, even without some of these advanced modules, remains competitive with or superior to prior works. YOLO-based baselines and state-of-the art methods, while strong, are outperformed by our method, especially on the challenging OpenUrine dataset. These results demonstrate the effectiveness of our architectural innovations for improving the automated analysis of urine sediment images.

Impact of input size

As shown in Table 6, an input resolution of 960×960 pixels yielded the highest overall mAP while maintaining stable convergence and feasible GPU memory usage (24 GB). Hence, this resolution was adopted for all subsequent experiments. The model achieves optimal results at 960×960 pixels, outperforming both smaller and even larger input sizes on almost all metrics. While a further increase to 1280×1280 yields competitive results, there is no consistent improvement and some metrics are slightly reduced, likely due to increased computational noise, overfitting, or diminished returns with upscaling. Notably, reducing the input size below 960 sharply degrades performance, especially for mAP50 and recall. This is particularly important because many urinary particles (such as bacteria and crystals) are small and easily lost at lower resolutions. At the lowest tested sizes (80 and 40 pixels), model recall especially collapses, confirming that sufficient image resolution is critical for the reliable detection of fine and small-scale particles. These findings underscore the need to optimize input size for automatic urine sediment analysis, balancing computational efficiency with the necessity to preserve particle detail.

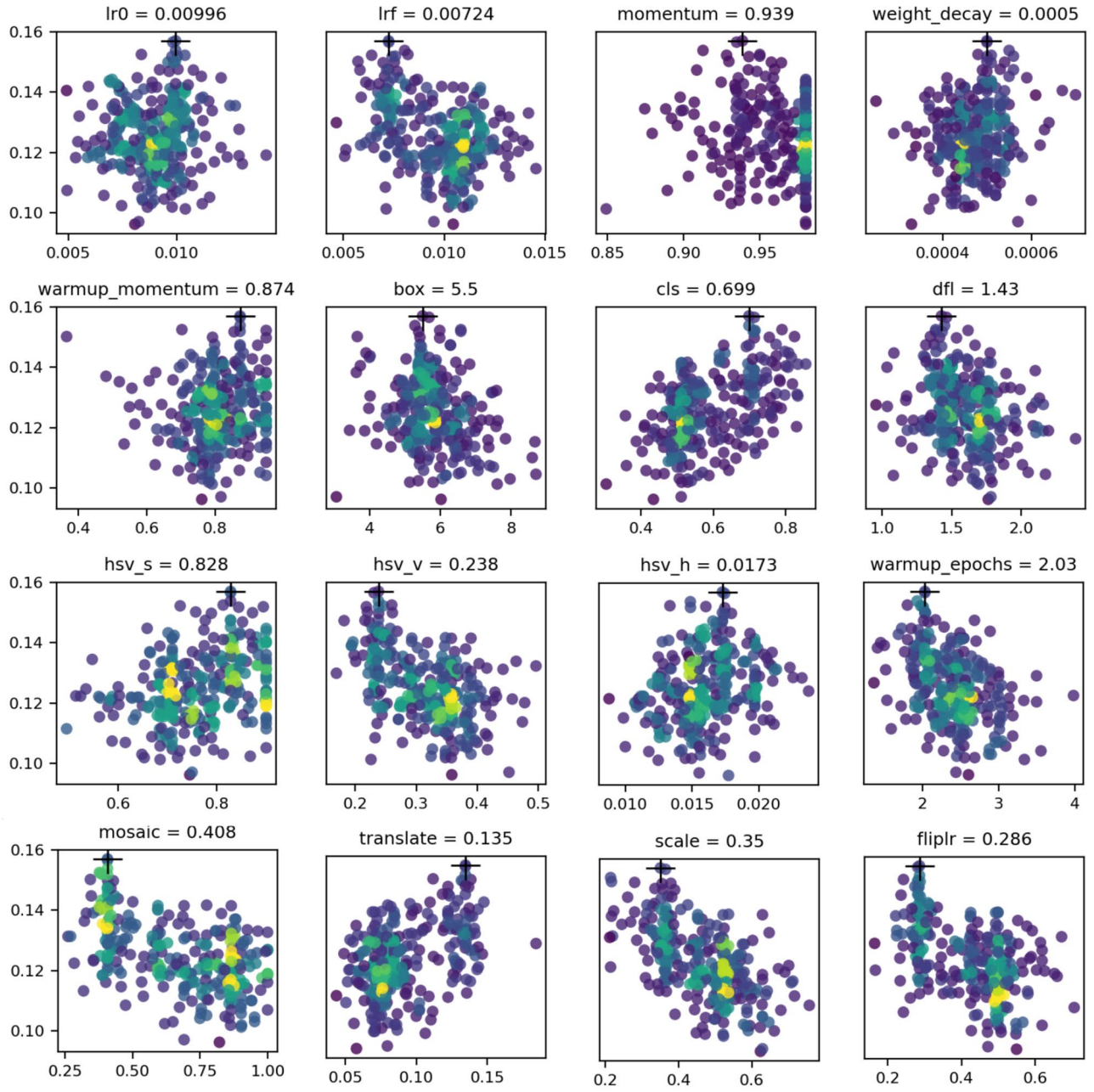

Fig. 4. Scatter plots illustrating the effect of key hyperparameters on object detection performance across 300 training runs (each for 100 epochs) on the OpenUrine dataset. Cross markers denote the final optimal values used for the main model training.

Impact of heads

The influence of each detection head was evaluated through a head-wise ablation test, as summarized in Table 5. When any single head was removed, its corresponding samples were not excluded from training; instead, all annotations were reassigned to the Others head to preserve the dataset composition and training balance. The results show that disabling any individual head consistently reduced detection accuracy, indicating that each contributes unique and complementary information. The most pronounced performance drop occurred when the Microorganisms head was removed, reflecting their critical role in discriminating morphologically complex or clinically significant particle groups.

Class-wise analysis

The results in Table 7 demonstrate that our method substantially improves detection accuracy across most urinary sediment particle classes compared to previous baselines. The model achieves high precision and recall on classes with distinctive morphological features, such as Calcium Oxalate, Bilirubin, and Calcium Carbonate, showing

Method	Precision (%)	Recall (%)	mAP50 (%)	mAP50-95 (%)
YOLO8-medium	57.85±2.00	54.43±0.84	57.39±0.30	41.66±1.03
YOLO9-medium	57.45±5.16	52.17±1.40	53.56±0.04	37.60±0.62
YOLO11-medium	60.27±0.52	54.81±1.19	58.30±0.22	43.10±0.22
YOLO12-nano	63.40±0.30	53.00±2.10	55.98±0.31	38.91±0.68
YOLO12-small	56.09±5.19	53.72±2.21	56.20±0.18	40.33±0.13
YOLO12-large	58.81±5.86	54.42±2.05	58.46±0.11	42.13±0.58
YOLO12-XLarge	58.51±0.91	57.06±3.69	58.40±0.24	41.63±1.00
EfficientDet-D2[55]	57.16±4.29	50.49±3.19	52.54±0.86	37.13±0.43
ViTDet[56]	56.67±2.31	52.70±3.81	54.90±1.18	38.21±0.89
RT-DETR large[57]	58.69±2.92	53.62±2.83	57.73±1.26	40.35±0.68
RT-DETRv2 large[58]	61.99±3.17	51.65±3.15	57.44±0.86	41.20±0.38
Liang et al. [23]	57.06±7.08	49.93±2.49	52.03±0.53	36.00±0.63
Liang et al. [29]	59.44±3.80	49.05±2.66	52.41±0.55	38.03±0.45
Li et al. [27]	62.79±4.15	50.30±2.02	52.89±0.20	36.81±0.26
Avci et al. [34]	55.16±5.24	54.53±2.97	53.09±0.08	36.58±0.32
Lyu et al. [7]	56.43±3.83	52.10±1.67	54.32±0.29	39.43±0.43
Wang et al. [59]	65.28±2.73	51.73±2.90	55.04±0.21	39.04±0.18
Komori et al. [60]	56.04±2.63	52.76±3.21	55.41±0.37	37.80±0.13
Suahil et al. [61]	63.86±2.94	51.21±0.86	55.77±0.12	40.68±0.40
Naznine et al. [62]	55.89±2.65	56.02±1.64	57.00±0.40	40.95±0.30
Akhtar et al. [63]	68.23±3.85	54.04±2.03	58.22±0.23	43.03±0.46
YOLOv12-M	62.45±3.35	54.54±2.28	58.66±0.08	41.88±0.65
YOLOv12-M + SAHI	62.91±2.68	56.49±1.98	59.91±0.74	42.27±0.80
YOLOv12-M + SAHI + Self-supervised + Single-head	63.20±5.13	58.60±2.30	61.08±0.36	44.44±0.74
YOLOv12-M + SAHI + Multi-head	70.19±5.20	60.34±2.83	62.01±0.49	45.21±0.75
Our method (YOLOv5-M backbone)	70.14±3.56	58.73±2.37	58.84±0.47	43.26±1.01
Our method (YOLOv8-M backbone)	72.12±4.37	60.36±3.71	59.04±0.52	44.79±0.29
Our method (YOLOv12-M backbone)	**76.59±3.29**	**62.45±2.94**	**64.15±1.56**	**47.20±0.72**

Table 4. Comparison of our proposed method and state-of-the-art methods on the OpenUrine dataset. Metrics are reported as mean±std (%) across five cross-validation folds. Bold values indicate the best result per metric.

Configuration	Precision (%)	Recall (%)	mAP@50 (%)	mAP@50:95 (%)
Without casts head	72.58±2.67	60.68±2.21	62.73±1.23	45.42±0.83
Without cells head	73.69±3.09	60.28±2.74	61.98±1.04	46.56±0.98
Without crystals head	73.58±2.39	61.71±2.39	60.38±1.13	44.70±0.87
Without microorganisms head	72.69±2.42	59.35±3.29	59.47±1.12	44.52±0.78
Full (all six heads)	76.59±3.29	62.45±2.94	64.15±1.56	47.20±0.72

Table 5. Head-wise ablation analysis of the proposed method. Removing each head degrades detection accuracy even though all data remain in use (annotations of the removed head were redirected to the others head).

the benefit of leveraging their unique visual patterns. However, certain classes remain challenging: for example, Bacteria achieve high precision but low recall, likely due to their small size and tendency to be overlooked in crowded backgrounds. Morphologically similar cells, notably RBC and WBC, are sometimes confused due to their overlapping appearance, limiting further accuracy improvements in these categories. Additionally, rare or subtle classes such as Fat Droplets and Renal Epithelial Cells still suffer from lower detection rates.

Typical failure cases include missed detections in densely clustered regions, merged bounding boxes where adjacent particles overlap, and occasional confusion between morphologically similar RBC and WBC, especially when illumination or focus artifacts blur their boundaries. Quantitative analysis indicates that approximately 38% of the undetected WBC instances were misclassified as RBC, while 43% of the missed RBC instances were incorrectly detected as WBC. This bidirectional confusion highlights their strong morphological resemblance under bright-field microscopy. In crowded microscopic fields, small Bacteria are sometimes undetected or

Input size	Precision (%)	Recall (%)	mAP50 (%)	mAP50-95 (%)
40	79.60±10.1	7.75±1.74	9.73±0.15	5.40±0.18
80	43.86±3.82	20.41±1.80	15.07±0.18	9.51±0.54
160	54.90±6.82	29.39±0.84	28.94±0.56	19.83±0.15
320	53.70±0.76	38.09±1.44	39.51±0.30	27.52±0.47
640	60.55±1.69	51.57±0.63	51.88±0.07	36.92±0.41
960	**63.40±0.30**	**53.00±2.10**	**55.98±0.31**	**38.91±0.68**
1280	56.29±3.97	53.49±3.08	54.00±0.30	36.44±0.45

Table 6. Evaluation results of YOLO12-Nano with different input image sizes on the OpenUrine test set. Metrics are reported as mean±std (%) across five cross-validation folds. The table demonstrates how increasing input resolution substantially improves the detection of urinary sediment particles, with the best performance achieved at 960×960 pixels.

merged with noise, while low-contrast Renal Epithelial Cells may be mistaken for background structures. These qualitative observations (illustrated in Fig.5) reveal the key limitations of the current model and inform future improvements such as boundary-aware loss design and targeted synthetic data augmentation for rare or visually ambiguous categories.

Overall, while our model demonstrates meaningful advances in most categories, the reliable detection of small, ambiguous, or visually similar particles remains a significant challenge for automated urine sediment analysis.

Clinical validation

Urine microscopy results from 84 patients, previously analyzed and verified by experienced laboratory technologists, were employed for clinical validation of the proposed model. For each sample, three to five representative microscopic fields were processed by the model, and the predictions were averaged at the patient level before comparison with the laboratory-reported results. The predicted outputs were mapped to the standard five-level microscopic quantitation scale (none, rare, few, moderate, many) used in routine clinical reporting. A prediction was considered correct when the model's categorical output matched the laboratory category for the same urinary component.

All major urinary sediment components, including RBCs, WBCs, epithelial cells, calcium oxalate crystals, bacteria, and mucus, were evaluated accordingly. Table 8 presents the clinical accuracy of the proposed model relative to technologist reports. Discrepant samples were further reviewed by an independent clinical biochemist to confirm the final reference label.

Interpretability via visual explanation

In Fig. 5, a comparison is presented between the model's bounding box predictions (left) and Grad-CAM visualizations (right) for selected test images. The detection results illustrate the network's ability to localize and classify different urine sediment constituents, such as amorphous particles and epithelial cells. Notably, the Grad-CAM activation maps reveal that the highlighted regions (red areas) are primarily concentrated over clear and well-defined particles within the microscopic fields, confirming that the model bases its predictions on relevant visual cues rather than background artifacts. This qualitative interpretability analysis demonstrates the reliability and transparency of the network's decision-making process in real-world clinical samples.

Conclusion

In this study, we introduced a novel deep learning method tailored for automated urine sediment analysis, integrating a multi-head YOLOv12 architecture, self-supervised pretraining, and SAHI-based inference. Our method effectively addresses critical challenges such as small-object detection, class imbalance, and data scarcity, leading to a competitive precision of 76.59% on a large, diverse dataset. The deployment of six specialized detection heads allows for detailed and simultaneous classification across all relevant urinary particles and artifacts, supporting detailed clinical interpretation. Furthermore, the establishment and public release of the OpenUrine dataset fill a crucial gap, providing a valuable resource for further research in this domain.

Future work will focus on refining the model's performance, especially for rare or visually ambiguous particle types, by exploring adaptive focal loss weighting, targeted synthetic data augmentation, and self-supervised consistency regularization to mitigate class imbalance. We also intend to integrate physical and chemical urinalysis test data to further enhance diagnostic precision and generalizability.

Class	Precision (%)	Recall (%)	mAP50 (%)	mAP50-95 (%)
Ammonium Biurat	100	68.88	84.28	57.96
Amorphous	82.60	85.40	85.54	47.46
Artifact	81.20	72.80	83.02	62.86
Bacteria	66.78	11.34	26.18	11.90
Bilirubin	100	100	100	100
Calcium Carbonate	100	100	100	100
Calcium Oxalate	95.76	90.16	98.56	65.66
Calcium Phosphate	51.10	70.00	70.56	53.34
Cholesterol	100	63.70	90.58	50.40
Cystine	94.50	35.00	54.46	42.42
Enterobius Vermicularis Egg	100	100	100	100
Epithelial Cell (Renal)	13.30	38.22	14.84	11.06
Epithelial Cell (Squamous)	98.28	100	100	74.34
Epithelial Cell (Transitional)	72.94	90.02	70.28	55.86
Fat Droplets	99.12	7.42	13.02	8.68
Fungal Hyphae	86.24	23.38	37.10	19.18
Granular Cast	100	49.98	73.50	45.22
Hippuric Acid	96.74	46.20	48.86	21.56
Hyaline Cast	85.12	70.00	75.18	36.82
Leucine	74.76	100	100	100
Lipid Cast	36.26	84.00	41.16	27.72
Mixed Cell Cast	100	98.00	100	100
Mucus	94.93	100	100	76.64
Muddy Brown Cast	100	73.36	98.84	69.30
Oval Fat Bodies	96.46	41.72	61.32	42.28
RBC	82.88	85.68	83.30	50.68
RBC Cast	62.02	56.00	67.62	55.02
RBC Clump	65.38	35.00	38.92	28.56
Schistosoma haematobium Eggs	43.96	100	100	100
Sperm	75.88	69.02	63.84	30.94
Suspected atypical cell	44.24	98.00	57.82	41.72
Trichomonas vaginalis	100	100	100	84.98
Triple Phosphate	92.26	94.50	95.62	64.54
Tyrosine crystal	33.74	38.22	28.42	14.00
Uric Acid	79.24	81.48	78.26	47.32
WBC	83.02	91.42	81.62	53.06
WBC Clump	43.12	3.64	5.18	3.64
Waxy Cast	100	52.50	59.36	49.70
Yeast	52.08	15.96	15.82	5.75
Unweighted average	**79.07**	**67.97**	**69.31**	**52.40**
Weighted average	**76.59**	**61.22**	**62.88**	**45.70**

Table 7. Class-wise detection performance of the best model on the OpenUrine test set. Numbers are reported as percentages. Metrics include Precision, Recall, mAP@50, and mAP@50-95 for each urinary sediment class, as well as unweighted and instance-weighted averages across all classes.

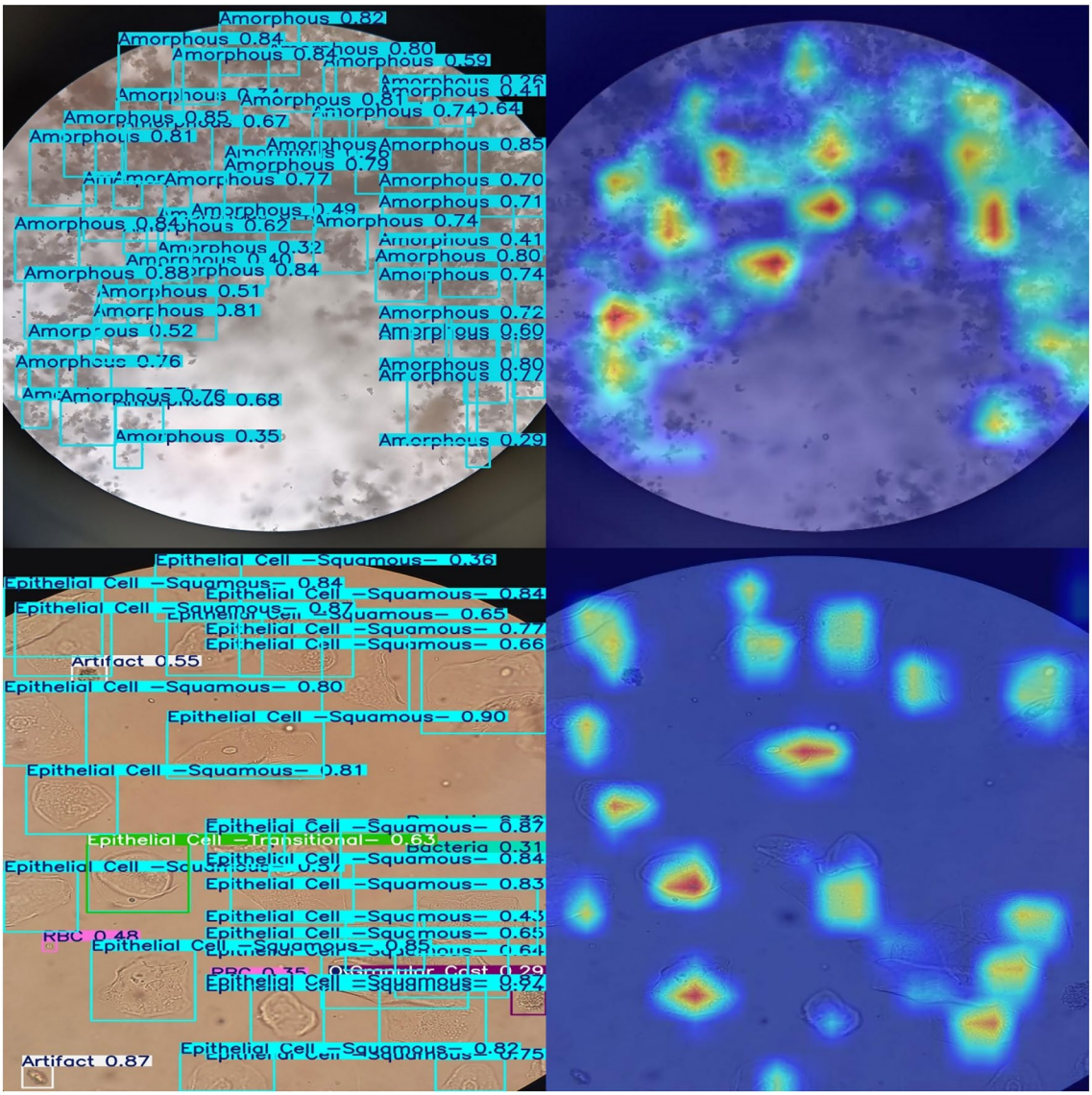

Fig. 5. Grad-CAM visualization of model attention across representative classes. Top row: detection and attention patterns for Amorphous particles, showing that the model accurately localizes dense crystalline regions and focuses its activations (red/yellow) on texture-rich clusters relevant to this class. Bottom row: predictions for Epithelial Cells where the model highlights cell nuclei and boundary contours while de-emphasizing background noise and staining artifacts. In each pair, the left image displays predicted bounding boxes and class labels, while the right image presents the corresponding Grad-CAM heatmap. Warmer colors (red/yellow) indicate regions contributing most to the network's decision, confirming that it primarily attends to morphologically informative structures such as epithelial cells and amorphous deposits rather than irrelevant background patterns.

Urinary component	Clinical accuracy (%)
RBC	87.31
WBC	88.40
Epithelial cells	97.10
Calcium oxalate	93.79
Bacteria	59.36
Mucus	93.32
Mean accuracy	86.55

Table 8. Patient-level accuracy of the proposed method for major urinary particle types compared with laboratory technologist reports.

Data availability

The OpenUrine dataset is available from the corresponding author on reasonable request. Interested researchers are invited to submit an application through the designated request form available at www.github.com/alikarimi120/OpenUrine. It should be noted that this dataset is provided solely for academic and non-commercial research purposes. Prior to access, requestors are required to agree to the terms and conditions specified in the request form, ensuring the data will be used in accordance with ethical standards and regulations.

Code availability

Sample codes, experiments results, and models are hosted on the following GitHub repository: www.github.com/alikarimi120/OpenUrine.

Received: 13 August 2025; Accepted: 20 October 2025
Published online: 21 November 2025

References

1. Simerville, J. A., Maxted, W. C. & Pahira, J. J. Urinalysis: A comprehensive review. *Am. Fam. Phys.* **71**, 1153–1162 (2005).
2. Cavanaugh, C. & Perazella, M. A. Urine sediment examination in the diagnosis and management of kidney disease: Core curriculum 2019. *Am. J. Kidney Dis.* **73**, 258–272 (2019).
3. Kouri, T. et al. European urinalysis guidelines. *Scand. J. Clin. Lab. Invest.* **60**, 1–96 (2000).
4. Winkel, P., Statland, B. & Joergensen, K. Urine microscopy, an iii-defined method, examined by a multifactorial technique. *Clin. Chem.* **20**, 436–439 (1974).
5. Yildirim, M., Bingol, H., Cengil, E., Aslan, S. & Baykara, M. Automatic classification of particles in the urine sediment test with the developed artificial intelligence-based hybrid model. *Diagnostics* **13**, 1299 (2023).
6. Xu, X.-T., Zhang, J., Chen, P., Wang, B. & Xia, Y. Urine sediment detection based on deep learning. In *Intelligent Computing Theories and Application: 15th International Conference, ICIC 2019, Nanchang, China, August 3–6, 2019, Proceedings, Part I 15*. 543–552 (Springer, 2019).
7. Lyu, H. et al. Automated detection of multi-class urinary sediment particles: An accurate deep learning approach. *Biocybern. Biomed. Eng.* **43**, 672–683 (2023).
8. Erten, M. et al. Automated urine cell image classification model using chaotic mixer deep feature extraction. *J. Digit. Imaging* **36**, 1675–1686 (2023).
9. Shen, D., Wu, G. & Suk, H.-I. Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.* **19**, 221–248 (2017).
10. Li, M., Jiang, Y., Zhang, Y. & Zhu, H. Medical image analysis using deep learning algorithms. *Front. Public Health* **11**, 1273253 (2023).
11. Ker, J., Wang, L., Rao, J. & Lim, T. Deep learning applications in medical image analysis. *IEEE Access* **6**, 9375–9389 (2017).
12. Cai, L., Gao, J. & Zhao, D. A review of the application of deep learning in medical image classification and segmentation. *Ann. Transl. Med.* **8**, 713 (2020).
13. Chen, T., Kornblith, S., Norouzi, M. & Hinton, G. A simple framework for contrastive learning of visual representations. In *International Conference on Machine Learning*. 1597–1607 (PmLR, 2020).
14. Huang, S.-C. et al. Self-supervised learning for medical image classification: a systematic review and implementation guidelines. *NPJ Digit. Med.* **6**, 74 (2023).
15. Ji, Q., Jiang, Y., Wu, Z., Liu, Q. & Qu, L. An image recognition method for urine sediment based on semi-supervised learning. *IRBM* **44**, 100739 (2023).
16. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **25** (2012).
17. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 770–778 (2016).
18. Szegedy, C. et al. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1–9 (2015).
19. Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4700–4708 (2017).
20. Howard, A. G. et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861 (2017).
21. Iriya, R. et al. Deep learning-based culture-free bacteria detection in urine using large-volume microscopy. *Biosensors* **14**, 89 (2024).
22. Uryu, H. et al. Automated urinary sediment detection for Fabry disease using deep-learning algorithms. *Mol. Genet. Metab. Rep.* **33**, 100921 (2022).
23. Liang, Y., Tang, Z., Yan, M. & Liu, J. Object detection based on deep learning for urine sediment examination. *Biocybern. Biomed. Eng.* **38**, 661–670 (2018).
24. Khalid, Z. M., Hawezi, R. S. & Amin, S. R. M. Urine sediment analysis by using convolution neural network. In *2022 8th International Engineering Conference on Sustainable Technology and Development (IEC)*. 173–178 (IEEE, 2022).

25. Girshick, R., Donahue, J., Darrell, T. & Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 580–587 (2014).
26. Liu, W. et al. SSD: Single shot multibox detector. In *European Conference on Computer Vision*. 21–37 (Springer, 2016).
27. Li, Q. et al. Inspection of visible components in urine based on deep learning. *Med. Phys.* **47**, 2937–2949 (2020).
28. Lin, T.-Y., Goyal, P., Girshick, R., He, K. & Dollár, P. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision*. 2980–2988 (2017).
29. Liang, Y., Kang, R., Lian, C. & Mao, Y. An end-to-end system for automatic urinary particle recognition with convolutional neural network. *J. Med. Syst.* **42**, 1–14 (2018).
30. Li, T. et al. The image-based analysis and classification of urine sediments using a LENET-5 neural network. *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.* **8**, 109–114 (2020).
31. LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2324 (2002).
32. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014).
33. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2818–2826 (2016).
34. Avci, D. et al. A new super resolution faster r-cnn model based detection and classification of urine sediments. *Biocybern. Biomed. Eng.* **43**, 58–68 (2023).
35. Avci, D., Leblebicioglu, M. K., Poyraz, M. & Dogantekin, E. A new method based on adaptive discrete wavelet entropy energy and neural network classifier (adweenn) for recognition of urine cells from microscopic images independent of rotation and scaling. *J. Med. Syst.* **38**, 1–9 (2014).
36. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn.* **20**, 273–297 (1995).
37. Erten, M. et al. Swin-lbp: A competitive feature engineering model for urine sediment classification. *Neural Comput. Appl.* **35**, 21621–21632 (2023).
38. Ji, Q., Li, X., Qu, Z. & Dai, C. Research on urine sediment images recognition based on deep learning. *IEEE Access* **7**, 166711–166720 (2019).
39. Nagai, T., Onodera, O. & Okuda, S. Deep learning classification of urinary sediment crystals with optimal parameter tuning. *Sci. Rep.* **12**, 21178 (2022).
40. Ge, Z., Liu, S., Wang, F., Li, Z. & Sun, J. Yolox: Exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430 (2021).
41. Hussain, T., Shouno, H., Mohammed, M. A., Marhoon, H. A. & Alam, T. DCSSGA-UNET: Biomedical image segmentation with DenseNet channel spatial and semantic guidance attention. *Knowl.-Based Syst.* **314**, 113233 (2025).
42. Hussain, T. et al. Effresnet-vit: A fusion-based convolutional and vision transformer model for explainable medical image classification. *IEEE Access* (2025).
43. Ali, Z. et al. Deep learning-driven cyber attack detection framework in dc shipboard microgrids system for enhancing maritime transportation security. In *IEEE Transactions on Intelligent Transportation Systems* (2025).
44. Ali, Z. et al. A novel hybrid signal processing based deep learning method for cyber-physical resilient harbor integrated shipboard microgrids. In *IEEE Transactions on Industry Applications* (2025).
45. Ali, Z. et al. A novel intelligent intrusion detection and prevention framework for shore-ship hybrid ac/dc microgrids under power quality disturbances. In *2025 IEEE Industry Applications Society Annual Meeting (IAS)*. 1–7 (IEEE, 2025).
46. Franzò, M., Pica, A., Pascucci, S., Marinozzi, F. & Bini, F. Hybrid system mixed reality and marker-less motion tracking for sports rehabilitation of martial arts athletes. *Appl. Sci.* **13**, 2587 (2023).
47. Ferraz, A., Duarte-Mendes, P., Sarmento, H., Valente-Dos-Santos, J. & Travassos, B. Tracking devices and physical performance analysis in team sports: a comprehensive framework for research—trends and future directions. *Front. Sports Active Living* **5**, 1284086 (2023).
48. Tian, Y., Ye, Q. & Doermann, D. Yolov12: Attention-centric real-time object detectors. arXiv preprint arXiv:2502.12524 (2025).
49. Tian, Y., Ye, Q. & Doermann, D. Yolov12: Attention-centric real-time object detectors (2025).
50. Akyon, F. C., Altinuc, S. O. & Temizel, A. Slicing aided hyper inference and fine-tuning for small object detection. In *2022 IEEE International Conference on Image Processing (ICIP)*. 966–970. https://doi.org/10.1109/ICIP46576.2022.9897990 (2022).
51. Zheng, Z. et al. Distance-IOU loss: Faster and better learning for bounding box regression. *Proc. AAAI Conf. Artif. Intell.* **34**, 12993–13000 (2020).
52. Li, X. et al. Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection. *Adv. Neural Inf. Process. Syst.* **33**, 21002–21012 (2020).
53. Bishop, C. M. & Nasrabadi, N. M. *Pattern Recognition and Machine Learning*. Vol. 4 (Springer, 2006).
54. Lin, T.-Y. et al. Microsoft coco: Common objects in context. In *European Conference on Computer Vision*. 740–755 (Springer, 2014).
55. Tan, M., Pang, R. & Le, Q. V. Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10781–10790 (2020).
56. Li, Y., Mao, H., Girshick, R. & He, K. Exploring plain vision transformer backbones for object detection. In *European Conference on Computer Vision*. 280–296 (Springer, 2022).
57. Zhao, Y. et al. Detrs beat yolos on real-time object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 16965–16974 (2024).
58. Lv, W. et al. Rt-detrv2: Improved baseline with bag-of-freebies for real-time detection transformer. arXiv preprint arXiv:2407.17140 (2024).
59. Wang, C. et al. Dp-yolo: Effective improvement based on yolo detector. *Appl. Sci.* **13**, 11676 (2023).
60. Komori, T., Nishikawa, H., Taniguchi, I. & Onoye, T. Improvement of yolov7 with attention modules for urinary sediment particle detection. In *2023 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. 1–5 (IEEE, 2023).
61. Suhail, K. & Brindha, D. Microscopic urinary particle detection by different yolov5 models with evolutionary genetic algorithm based hyperparameter optimization. *Comput. Biol. Med.* **169**, 107895 (2024).
62. Naznine, M., Salam, A., Khan, M. M., Nobi, S. F. & Chowdhury, M. E. An ensemble deep learning approach for accurate urinary sediment detection using yolov9e and kd-yolox-vit. *IEEE Access* (2025).
63. Akhtar, S., Hanif, M., Saraoglu, H. M., Lal, S. & Arshad, M. W. Yolov11-samnet: A hybrid detection and segmentation framework for urine sediment analysis. In *International Conference on Computational Science and Its Applications*. 241–251 (Springer, 2025).

Acknowledgements

The authors acknowledge the use of artificial intelligence-assisted tools, including ChatGPT (OpenAI) and Grammarly, for improving the clarity, grammar, and overall readability of the manuscript. The authors are solely responsible for the content and interpretation of the findings presented in this paper. A part of the article has been extracted from the thesis written by Mehdi Alizadeh in the School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran (registration number: 43011137). The local Committee for Ethics approved the study (reference number: IR.SBMU.MSP.REC.1403.318).

Author contributions

M.A. collected the data, contributed to unlabeled data annotation, and assisted in manuscript preparation. A.K. performed data preprocessing, wrote the manuscript, designed and conducted the experiments, and developed the proposed method. M.J.B. contributed to the implementation of the proposed method. A.M. and S.A supervised the project. M.S. revised the manuscript and supervised the project. M.A.A. verified the theoretical findings, revised the manuscript, and supervised the project. All authors reviewed and approved the final manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Ethical approval

In accordance with ethical guidelines, this study involving human participants was reviewed and approved by the Ethics Committee of Shahid Beheshti University of Medical Sciences (approval code: IR.SBMU.MSP.REC.1403.318). Prior to participation, all individuals provided written informed consent. Additionally, explicit consent for the publication of any images included in this manuscript was obtained from each participant.

Additional information

Correspondence and requests for materials should be addressed to M.S.-S. or M.A.A.