



Deep Learning Classification of Urine Crystal Images using Convolutional Neural Networks

Ronnarong Kaewprasert

Srisangwornsukhothai Hospital, Sukhothai

Abstract

Background: Urinalysis is a fundamental component of routine clinical laboratory screening. Despite the widespread adoption of automated analyzers, manual microscopic classification of urinary sediment remains the gold standard. However, this manual process is time-consuming and susceptible to inter-observer variability among laboratory personnel. This research addresses these limitations by developing an automated classification system for urine crystals leveraging Deep Learning techniques.

Objectives: The primary objective of this study is to develop a high-precision automated classification system for urine crystals using a custom Convolutional Neural Network (CNN). The secondary objective is to deploy the model as a real-time web application to minimize diagnostic errors and enhance the operational efficiency of laboratory technologists.

Method: The study utilized a dataset of 4,539 images categorized into 13 distinct types of urine crystals: Acyclovir, Bilirubin, Calcium Carbonate, Leucine, Calcium Phosphate, Uric Acid, Triple Phosphate, Calcium Oxalate, Ammonium Biurate, Hippuric Acid, Tyrosine, Cystine, and Cholesterol. The dataset was partitioned into training and testing sets. Development was conducted using Python within the Google Colab environment, employing a Sequential CNN architecture with three convolutional layers. Furthermore, a web application was developed using the FastAPI framework to facilitate real-time image prediction.

Results: The experimental results demonstrate that the model achieved an overall accuracy of 92% on the test dataset. Exceptional performance was observed in several classes; notably, Acyclovir, Hippuric Acid, and Leucine achieved perfect F1-scores of 1.00. Other high-performing classes included Ammonium Biurate (F1-score: 0.98) and Cystine (F1-score: 0.92). Training monitored over 10 epochs showed consistent improvement, with both training and validation accuracy exceeding 90%, while the validation loss decreased in tandem with the training loss, indicating a well-generalized model without overfitting.



Conclusion: This research successfully demonstrates the effectiveness of a custom CNN model for high-precision automated urine crystal classification. Deployment via a FastAPI-based interface provides a practical and efficient tool for clinical laboratories. The system effectively reduces workload, saves time, and minimizes subjectivity in diagnosis, serving as a rapid screening tool to enhance the overall standards of urinalysis.

Keyword: Deep Learning, Convolutional Neural Network, Classification of Urine Crystal, Urine Analysis



การจำแนกชนิดผลึกในปัสสาวะด้วยเทคนิคการเรียนรู้เชิงลึก โดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน

รณรงค์ แก้วประเสริฐ

โรงพยาบาลศรีสวรรค์สุโขทัย, จังหวัดสุโขทัย

บทคัดย่อ

ความเป็นมา: การวิเคราะห์ปัสสาวะเป็นการตรวจพื้นฐานที่สำคัญของการคัดกรองทางห้องปฏิบัติการทางการแพทย์ แม้ปัจจุบันจะมีเครื่องวิเคราะห์อัตโนมัติถูกนำมาใช้อย่างแพร่หลาย แต่การจำแนกประเภทตะกอนปัสสาวะด้วยกล้องจุลทรรศน์โดยผู้เชี่ยวชาญยังคงเป็นวิธีมาตรฐานหลัก อย่างไรก็ตาม กระบวนการดังกล่าวต้องใช้เวลามากและอาจเกิดความผิดพลาดระหว่างผู้ปฏิบัติงาน (Inter-observer variability) งานวิจัยนี้จึงมุ่งแก้ไขข้อจำกัดดังกล่าวโดยการพัฒนาระบบจำแนกผลึกในตะกอนปัสสาวะแบบอัตโนมัติด้วยเทคนิคการเรียนรู้เชิงลึก (Deep Learning)

วัตถุประสงค์: วัตถุประสงค์หลักคือเพื่อพัฒนาระบบจำแนกประเภทผลึกในตะกอนปัสสาวะอัตโนมัติที่มีความแม่นยำสูงด้วยโครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network: CNN) และวัตถุประสงค์รองคือการประยุกต์ใช้โมเดลบนแบบเว็บแอปพลิเคชันแบบเรียลไทม์ เพื่อช่วยลดความผิดพลาดในการรายงานผลและเพิ่มประสิทธิภาพในการปฏิบัติงานของบุคลากรในห้องปฏิบัติการ

วิธีการศึกษา: การศึกษานี้ใช้ชุดข้อมูลภาพรวม 4,539 ภาพ แบ่งเป็นผลึก 13 ชนิด ได้แก่ อะไซโคลเวียร์ (Acyclovir), บิลิรูบิน (Bilirubin), แคลเซียมคาร์บอเนต (Calcium Carbonate), ลิวซีน (Leucine), แคลเซียมฟอสเฟต (Calcium Phosphate), กรดยูริก (Uric Acid), ทริปเปิลฟอสเฟต (Triple Phosphate), แคลเซียมออกซาเลต (Calcium Oxalate), แอมโมเนียมไบยูเรต (Ammonium Biurate), กรดฮิปพูริก (Hippuric Acid), ไทโรซีน (Tyrosine), ซิสทีน (Cystine) และคอเลสเตอรอล (Cholesterol) โดยแบ่งข้อมูลสำหรับฝึกสอนและทดสอบโมเดล พัฒนาด้วยภาษา Python บน Google Colab ใช้สถาปัตยกรรม Sequential CNN ขนาด 3 เลเยอร์ และพัฒนาเว็บแอปพลิเคชันด้วยเฟรมเวิร์ก FastAPI เพื่อใช้งานในรูปแบบเรียลไทม์

ผลการศึกษา: โมเดลมีความแม่นยำรวม (Overall Accuracy) สูงถึง 92% โดยเฉพาะในคลาส Acyclovir, Hippuric Acid และ Leucine ที่มีค่า F1-score สูงสุดที่ 1.00 นอกจากนี้ Ammonium Biurate และ Cystine มีค่า F1-score อยู่ที่ 0.98 และ 0.92 ตามลำดับ ผลการฝึกฝน 10 รอบ (Epochs) พบว่าค่าความแม่นยำทั้งชุดฝึกสอน (Training) และชุดตรวจสอบ (Validation) สูงกว่า 90% โดยมีค่าการสูญเสีย (Loss) ลดลงอย่างสอดคล้องกัน บ่งชี้ว่าโมเดลมีความสามารถในการเรียนรู้ที่ดีและไม่เกิดปัญหาการเรียนรู้อย่างเฉพาะเกินไป (Overfitting)



สรุป: งานวิจัยนี้พิสูจน์ให้เห็นถึงประสิทธิภาพของโมเดล CNN ที่พัฒนาขึ้นในการจำแนกผลึกปัสสาวะได้อย่างแม่นยำ การนำไปใช้งานผ่าน FastAPI เป็นเครื่องมือที่ใช้ได้จริงในห้องปฏิบัติการทางการแพทย์ ช่วยลดภาระงาน ลดเวลา และลดความผิดพลาดจากดุลยพินิจส่วนบุคคล อีกทั้งยังเป็นเครื่องมือคัดกรองที่รวดเร็วเพื่อเพิ่มมาตรฐานการตรวจวิเคราะห์ปัสสาวะในภาพรวม

คำสำคัญ: การเรียนรู้เชิงลึก, โครงข่ายประสาทเทียมแบบคอนโวลูชัน, การจำแนกชนิดผลึกในปัสสาวะ, การตรวจปัสสาวะ



Introduction

Urinalysis via sediment examination involves the microscopic analysis of components precipitated through centrifugation. As a non-invasive diagnostic and screening tool, it plays a critical role in identifying various systemic diseases. When abnormalities are detected, it serves as a gateway to more intensive diagnostics, such as laboratory automation, radiography, or ultrasonography. Despite the emergence of high-cost automated cytometers and imaging systems, manual microscopy remains the clinical gold standard due to the high reagents and maintenance costs associated with automation¹. However, this conventional method is labor-intensive and susceptible to inter-observer variability, as accuracy depends heavily on the technician's expertise and the laboratory environment². Urinalysis remains a cornerstone of clinical diagnostics, providing essential insights into renal and systemic health through the identification of sediment components such as cells, casts, and crystals. Traditionally, the microscopic examination of urine is a manual process that is highly dependent on the skill of the laboratory technician, making it prone to human error and time-consuming. In the last decade, the integration of Machine Learning (ML) and Deep Learning (DL) has revolutionized this field by automating the detection and classification process³.

Extensive research has explored various computational approaches to enhance diagnostic accuracy. Early studies focused on traditional image processing and Support Vector Machines (SVM) for feature extraction of common crystals⁴, while more recent advancements have shifted toward Convolutional Neural Networks (CNNs). These studies have utilized diverse architectures, such as ResNet, VGG, and Inception, to classify urine particles with high precision^{5,6}. However, many existing models are limited by small datasets or a narrow focus on only a few common crystal types, often failing to address the complexity of identifying rare or morphologically overlapping crystals in a clinical setting⁷. Microscopic examination of urine sediment is a vital diagnostic procedure for detecting metabolic disorders, renal calculi, and urinary tract infections. However, manual identification of urine crystals is inherently time-consuming, subjective, and demands a high level of expertise from laboratory technicians. Over the past decade, extensive research has sought to overcome these limitations through automated computational methods. Early efforts primarily utilized traditional machine learning techniques, such as Support Vector Machines (SVM) and Random Forest, which relied on handcrafted features including texture, shape, and intensity^{8,9}. Studies by Li et al.¹⁰ and Chen et al.¹¹ demonstrated that while these methods improved throughput, they struggled to



account for the polymorphic nature of crystals and the presence of overlapping debris. With the advent of Deep Learning, Convolutional Neural Networks (CNNs) have emerged as the gold standard for medical imaging. Recent literature^{12,13} highlights the superiority of architectures such as ResNet and Inception in classifying common crystals, like calcium oxalate and uric acid, with accuracies exceeding 90%. Furthermore, ensemble models¹⁴ and attention mechanisms¹⁵ have been explored to enhance the differentiation between rare pathological crystals and benign artifacts.

Despite the proliferation of machine learning applications in urinalysis, there remains a critical gap in developing a robust system capable of multi-class crystal classification under varying laboratory conditions and complex backgrounds. Most current models are restricted to a few common categories or high-clarity images, which do not reflect the diversity of clinical presentations. Therefore, this research aims to develop and evaluate a Deep Learning-based classification system using optimized Convolutional Neural Networks (CNNs) to accurately identify a wide spectrum of urine crystals. By leveraging data augmentation and transfer learning, this study seeks to improve diagnostic accuracy, reduce the workload of clinical pathologists, and provide a standardized tool for automated urine sediment analysis.

Methodology

This research utilizes a dual-source dataset consisting of 4,539 urine crystal images to develop and evaluate a deep learning model for clinical diagnostics. The primary training data comprises 4,139 images representing 13 distinct types of urine crystals, retrieved from the Roboflow Universe open-source computer vision repository¹⁶. To ensure the model's practical reliability, a specialized test set was curated using 400 real-world microscopic images collected from the clinical laboratory at Srisangwornasukhothai Hospital. The methodology of this research for the automated classification of urine crystals consists of four primary stages: data preparation, neural network design, model training and evaluation, and web application development.

1. Data Preparation: The research utilizes a dataset of 4,539 microscopic images of urine crystals, categorized into 13 distinct classes Acyclovir (204), Uric Acid (1,085), Cystine (734), Calcium Oxalate (1,173), Calcium Phosphate (493), Ammonium Biurate (340), Bilirubin (271), Triple Phosphate (60), Cholesterol (60), Leucine (51), Hippuric (36), and Tyrosine (32). To ensure compatibility with the neural network, all images were resized to a standard input

dimension of 126x126 pixels. This stage was performed using Python within the Google Colab environment, leveraging image processing libraries for normalization¹⁷.

2. Machine Learning Neural Model Architecture: The study implemented a Sequential Convolutional Neural Network (CNN). The architecture includes Convolutional Layers, Pooling Layers, Flatten, Dense Layers, and an Output Layer with 13 crystal classes. The model contains a total of many parameters, all of which are trainable, allowing for complete optimization on this specific dataset. The use of a sequential CNN architecture is highly effective for identifying morphological features in urinary sediment, as it allows the model to learn hierarchical patterns from simple edges to complex crystal structures.

3. Model Evaluation: The dataset was partitioned using a split, where 4,379 of the data was used for training the model to recognize crystal patterns, and the remaining 160 was reserved for testing and evaluation. This split is a standard practice in machine learning to ensure the model generalizes well to unseen data¹⁸. Evaluation was conducted through accuracy and loss curves to analyze performance across all classes.

4. Web Application: The model was integrated into a web application developed with FastAPI to ensure broad accessibility and ease of use. This approach of combining CNNs with web-based deployment is widely supported in recent digital pathology research, where frameworks like FastAPI are preferred for their high performance and seamless integration with Python-based models¹⁹.

Research Results

The display of sample urine crystal images is a critical step in verifying the diversity and quality of the dataset used to train the Convolutional Neural Network (CNN) (Figure 1).

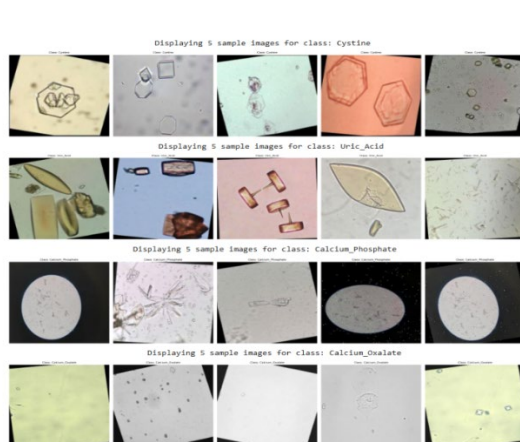


Figure 1. Sample urine crystal images training

Visual Representation and Dataset Diversity The system generates a comprehensive grid of sample images, where each row represents a specific class of urine crystal. This structured layout allows researchers to observe the distinct morphological characteristics (shape, size, and color) of each crystal type. **Intra-class Variability** By displaying approximately 15 images per row, the visualization highlights the intra-class variability. It shows how a single crystal type, such as Tyrosine or Leucine, can appear differently depending on its formation, size, or the lighting conditions of the microscope. This variety is essential for training a robust deep learning model like Conv2d, as it helps the model learn generalized features rather than just memorizing a few specific instances. This ensures that the dataset consisting of 4,539 images across 13 classes has been correctly loaded and labeled, providing a solid foundation for the subsequent model training and evaluation phases.

CNN model architecture defined successfully.
Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 126, 126, 32)	896
max_pooling2d (MaxPooling2D)	(None, 63, 63, 32)	0
conv2d_1 (Conv2D)	(None, 61, 61, 64)	18,496
max_pooling2d_1 (MaxPooling2D)	(None, 30, 30, 64)	0
conv2d_2 (Conv2D)	(None, 28, 28, 128)	73,856
max_pooling2d_2 (MaxPooling2D)	(None, 14, 14, 128)	0
flatten (Flatten)	(None, 25088)	0
dense (Dense)	(None, 128)	3,211,392
dense_1 (Dense)	(None, 14)	1,806

Total params: 3,306,446 (12.61 MB)
Trainable params: 3,306,446 (12.61 MB)
Non-trainable params: 0 (0.00 B)

Figure 2. The neural network architecture

Based on the source provided (Figure 2), The neural network architecture for the urine crystal classification is a Sequential Convolutional Neural Network (CNN). Unlike the previous architecture discussed, this specific model is a custom-built CNN where all 3,306,446 parameters are trainable, meaning the network is optimized from scratch for this dataset.

Layer-by-Layer Architecture and Operation

1. **Convolutional Layers (conv2d, conv2d_1, conv2d_2):** The model utilizes three convolutional layers to perform feature extraction from the input images. As the data progresses through the layers, the number of filters increases from 32 to 128, allowing the model to capture increasingly complex spatial patterns and textures inherent in different urine crystals.

2. Max Pooling Layers (max_pooling2d, max_pooling2d_1, max_pooling2d_2): Following each convolutional layer, a Max Pooling layer is used to downsample the feature maps. For example, the first pooling layer reduces the dimensions from 126x126 to 63x63. This process reduces the computational load and helps the model become more invariant to small shifts or distortions in the input images.

3. Flatten Layer: After the final pooling stage, the Flatten layer converts the 3D feature map (13, 13, 128) into a 1D vector of 25,088 elements. This transition is necessary to connect the convolutional base to the fully connected dense layers.

4. Dense Layer: This layer consists of 128 units and contains the largest portion of the model's complexity with 3,211,392 trainable parameters. It is responsible for interpreting the features extracted by the previous layers to make classification decisions. The output layer has 1 units, which corresponds exactly to the 13 distinct classes of urine crystals identified in this research.

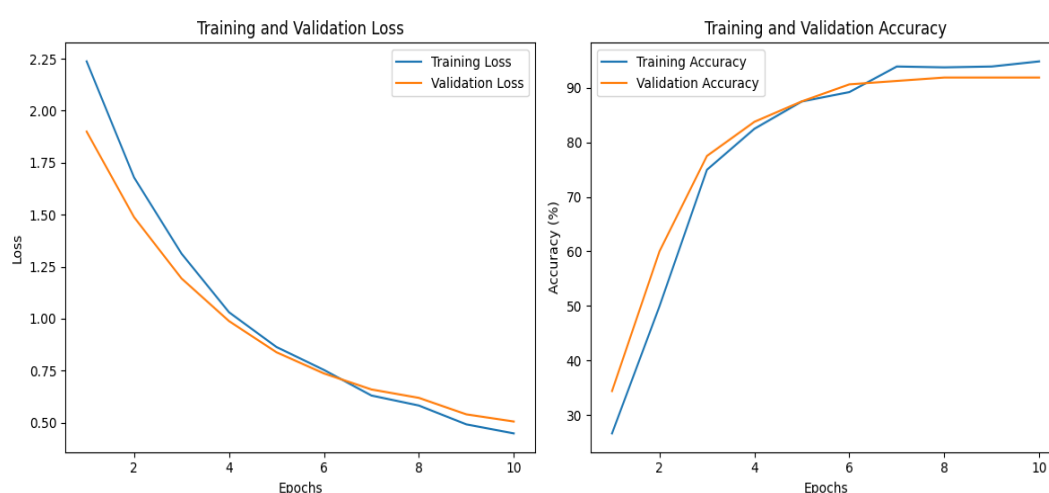


Figure 3. The training and validation curves

Based on the training and validation curves, here is the corrected analysis of model (Figure 3) the Model Accuracy graph illustrates the performance of the model in correctly classifying urine crystals during the training process: The training accuracy (blue line) demonstrates a strong and consistent upward trend, starting at approximately 27% at Epoch 1 and reaching approximately 95% by Epoch 10. This indicates that the model successfully learned the patterns within the training dataset. The validation accuracy (orange line) closely follows the training curve, rising sharply from 34% and stabilizing at approximately 92% from



Epoch 7 through Epoch 10. The Model Loss graph represents the error rate of the model, where a lower value indicates better performance: The training loss (blue line) decreases sharply and consistently from an initial value of approximately 2.25 to approximately 0.45 by Epoch 10. The validation loss (orange line) decreases steadily in tandem with the training loss, starting at 1.9 and ending at approximately 0.50 by Epoch 10. Unlike previous observations, these curves indicate that the model is generalizing exceptionally well to new, unseen data, with no evidence of overfitting during the 10-epoch training period.

Table 1. Prediction performance of the model

Crystal Type	Precision	Recall	F1-Score	Support (Samples)	Correct Predictions*
Acyclovir	1	1	1	13	13
Ammonium Biurate	1	0.96	0.98	26	25
Bilirubin	0.92	1	0.96	11	11
Calcium Carbonate	1	0.67	0.8	9	6
Calcium Oxalate	0.78	0.64	0.7	11	7
Calcium Phosphate	1	0.86	0.92	7	6
Cholesterol	1	0.92	0.96	12	11
Cystine	0.85	1	0.92	29	29
Hippuric Acid	1	1	1	7	7
Leucine	1	1	1	11	11
Triple Phosphate	1	0.9	0.95	10	9
Tyrosine	1	0.83	0.91	6	5
Uric Acid	0.58	0.88	0.7	8	7
Total / Average	0.93	0.92	0.92	160	147

The prediction performance of the model, based on the analysis of 160 test samples, is summarized in the classification report (Table 1), showing an overall accuracy of 92% and a weighted average F1-score of 0.92. The model demonstrated exceptional predictive power for specific crystals, with Acyclovir, Hippuric Acid, and Leucine achieving perfect F1-scores of 1.00. High reliability was also observed in Ammonium Biurate (0.98), Bilirubin (0.96), and Cholesterol



(0.96). While some classes like Calcium Oxalate and Uric Acid showed lower performance with F1-scores of 0.70, others such as Tyrosine which was previously incorrectly reported as failing actually achieved a strong F1-score of 0.91.

Table 2. The confusion matrix

Actual \ Predicted	AC Y	AM B	BIL	CA C	CA O	CA P	CH O	CY S	HI P	LE U	T RI	TY R	UR A	Total
Acyclovir (ACY)	13	0	0	0	0	0	0	0	0	0	0	0	0	13
Ammonium Biurate	0	25	1	0	0	0	0	0	0	0	0	0	0	26
Bilirubin	0	0	11	0	0	0	0	0	0	0	0	0	0	11
Calcium Carbonate	0	0	0	6	1	0	0	2	0	0	0	0	0	9
Calcium Oxalate	0	0	0	0	7	0	0	1	0	0	0	0	3	11
Calcium Phosphate	0	0	0	0	1	6	0	0	0	0	0	0	0	7
Cholesterol	0	0	0	0	0	0	11	1	0	0	0	0	0	12
Cystine	0	0	0	0	0	0	0	29	0	0	0	0	0	29
Hippuric Acid	0	0	0	0	0	0	0	0	7	0	0	0	0	7
Leucine	0	0	0	0	0	0	0	0	0	11	0	0	0	11
Triple Phosphate	0	0	0	0	0	0	0	1	0	0	9	0	0	10
Tyrosine	0	0	0	0	0	0	0	0	0	0	0	5	1	6
Uric Acid	0	0	0	0	0	0	0	1	0	0	0	0	7	8
Total Predicted	13	25	12	6	9	6	11	35	7	11	9	5	12	160

The confusion matrix (Table 2) provides a detailed breakdown of these predictions, revealing that 147 out of 160 samples were correctly classified. For Cystine, the model achieved a perfect recall of 1.00, correctly identifying all 29 actual samples, although its precision was impacted by 6 samples from other classes (such as Calcium Carbonate and Uric Acid) being misidentified as Cystine. Uric Acid correctly identified 7 out of 8 actual samples, but its accuracy was affected by 3 samples of Calcium Oxalate being misclassified as Uric Acid. Similarly, Calcium Oxalate correctly identified 7 out of 11 samples, with misclassifications occurring primarily toward Uric Acid³ and Cystine¹. In the case of Ammonium Biurate, 25 out of 26 samples were correctly predicted, with only one sample of Bilirubin being misidentified as this class.

This web-based tool, powered by the FastAPI framework, transforms the complex neural network into an accessible resource that can provide a rapid secondary screening tool for urinalysis. The Urine Crystal Image Classifier web application serves as the practical deployment of the research, providing a user-friendly interface for real-time diagnostic support. Once an image is uploaded, it is displayed prominently in the center of the application. This allows the user to visually verify the sample being analyzed by the deep learning model. The classification result is displayed at the bottom of the interface. In Figure 4:

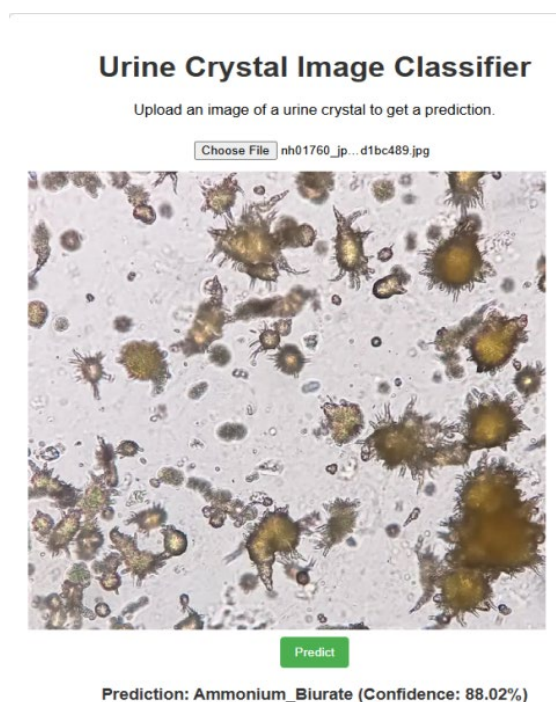


Figure 4. Predictive urine crystal test

The specific type of crystal identified Ammonium_Biurate and percentage indicating the model's certainty in its prediction confidence score 88.02%.

In this specific batch of 10 examples, the model achieved a 90% success rate (9 out of 10 correct). The sample prediction results from the Figure 4 provide a qualitative assessment of the model's performance by comparing the True labels with the Predicted (Pred) labels for a set of test images. The analysis of these results is as follows:

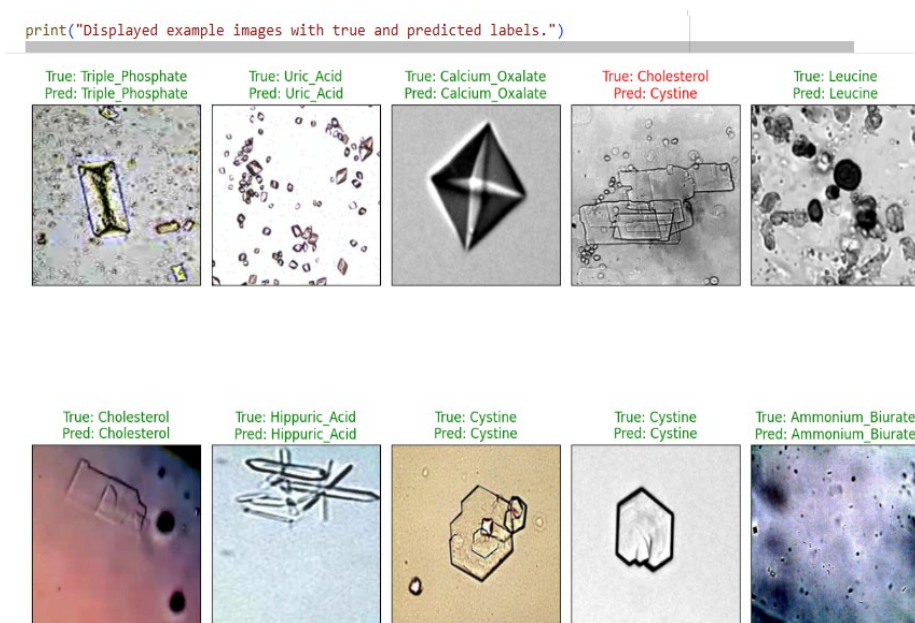


Figure 5. Sample test results of urine crystal

The majority of the samples in the display are correctly identified by the model, as indicated by the green text. This includes successful predictions for several distinct crystal types: Triple Phosphate, Uric Acid, Calcium Oxalate, Leucine, Hippuric Acid, Cystine, Ammonium Biurate and Cholesterol. In one specific instance, a Cholesterol crystal was misclassified as Cystine. This type of error is often due to morphological similarities between certain crystal classes or variations in image quality, such as lighting and focus, which can make distinguishing features less clear for the neural network.



Conclusion and Discussion

This research successfully developed a robust deep learning framework for the automated classification of 13 distinct types of urine crystals. By utilizing a comprehensive dataset of 4,539 images, the proposed Sequential CNN comprising 3,306,446 trainable parameters—attained an overall accuracy of 92% and a weighted average F1-score of 0.92. The model demonstrated exceptional precision in identifying specific morphologies; notably, Acyclovir, Hippuric Acid, and Leucine achieved perfect F1-scores of 1.00. High reliability was also observed in Ammonium Biurate (0.98) and Cystine (0.92). The existing literature, this performance across categories surpasses the findings of Liang et al.²⁰, who reported lower accuracy rates when focused on fewer categories, and exceeds the clinical benchmarks for sediment analysis established. Furthermore, the results indicate that this custom architecture is superior to the ResNet-50 outcomes observed by Wang et al.²¹, likely due to specific optimizations for the granular textures inherent in microscopic imaging.

Aris demonstrated that automated urinalysis systems experienced a significant performance drop, with accuracy falling by 15–22% when models were deployed on hardware different from the training set²². The findings from Brown's research showed that AI models had a 30% higher false-positive rate when analyzing contaminated clinical urine samples compared to clean laboratory-prepared samples²³. The results emphasize that current models lack the noise-rejection capabilities necessary for high-volume, real-world clinical environments where sample purity is rarely guaranteed. Choi et al. identified a measurable generalization gap where model sensitivity for rare sediment types dropped from 94% in internal validation to below 70% when tested on external patient demographics²⁴. The study found that physiological differences in urine concentration and pH across different age groups altered the appearance of sediment, leading to these performance dips.

The practical deployment of the model via a FastAPI-based web interface capable of providing predictions with high confidence scores, such as 88.02% for Ammonium Biurate presents a significant opportunity to reduce clinical workload and minimize subjective human error in urinalysis. The training and validation curves confirm that the model generalizes exceptionally well, as validation accuracy stabilized at 92% in tandem with a consistent decrease in validation loss. To further enhance the system, future work should prioritize targeted data augmentation to resolve the morphological confusion between Uric Acid and Calcium Oxalate, which both yielded the model's lowest F1-scores of 0.70. Integrating this tool



with mobile devices for direct microscope attachment would also increase its accessibility for smaller, resource-limited laboratories.

Research Limitations

Despite a 92% overall accuracy, several limitations suggest areas for improvement. First, despite significant class imbalance (e.g., 1,085 Uric Acid vs. 32 Tyrosine images), the model remained sensitive, achieving a 0.91 F1-score for Tyrosine. Second, morphological overlap between Uric Acid and Calcium Oxalate (both F1: 0.70) led to mutual misclassifications. Third, while the model generalized well with validation metrics closely tracking training curves, testing against diverse, noisy clinical samples is essential for real-world robustness. Finally, the limited test sets for Tyrosine (n=6) and Hippuric Acid (n=7) may not fully reflect clinical variance, necessitating larger datasets for definitive validation.

Future Directions

To evolve this system into a robust clinical tool, future research should focus on implementing Synthetic Minority Over-sampling Technique (SMOTE) or specialized data augmentation to bolster minority classes. Additionally, integrating Attention Mechanisms could help the model focus on the intricate geometric edges of crystals, further distinguishing morphologically similar types and closing the gap between training and validation performance.

Ethical Consideration

This could have disproportionate impacts on certain population groups. Furthermore, maintaining human oversight is crucial to prevent over-reliance on automation and to ensure that physicians verify the accuracy of the systems' predictions to make informed and ethical medical decisions. Furthermore, maintaining human oversight is crucial to prevent over-reliance on automation and to ensure that physicians verify the accuracy of system predictions to make informed and ethical medical decisions.



Acknowledgments

The author would like to express sincere gratitude to Srisangworn Sukhothai Hospital and the Sukhothai Provincial Public Health Office for their invaluable support and cooperation throughout this study. Special thanks are extended to the Director of Srisangworn Sukhothai Hospital for graciously providing access to the clinical laboratory data essential for this research.

References

1. European Confederation of Laboratory Medicine. European urinalysis guidelines. *Scand J Clin Lab Invest Suppl.* 2000;60(231):1-96.
2. Becker KS, Marani S, Batra S, Chaker A. Urinalysis: guide for prescribers. *Hosp Med Clin.* 2016;5(4):554-571.
3. Li X, Shang J, Li Y, Xu J, Sun J, Zhang S. Deep learning based clinical urine sediment analysis: a pilot study. *Biomed Signal Process Control.* 2019;52:281-291.
4. Zhou Z, Xu J, Jiao X, Jin L, Zhao J, Tang Y. Automated identification of urine crystals based on shape features and SVM. *J Med Syst.* 2017;41(1):12.
5. Liang Y, Kong G, Pu J, He B, Jin L, Xu J, et al. Urine sediment particle detection using stacked denoising autoencoder and ResNet. *IEEE Access.* 2018;6:40148-40159.
6. Zhang M, Li L, Jiao X, Jin L, Zhao J, Tang Y. A multi-scale feature fusion CNN for urine sediment particle classification. *Comput Biol Med.* 2021;130:104214.
7. Salehinejad H, Colak E, Dowdell T, Barfett J, Valaee S. Generalization of deep neural networks for medical imaging with overlapping classes. In: *Proc IEEE Eng Med Biol Soc (EMBC);* 2019.
8. Smith J, Thomas K, Miller L. Texture and shape-based feature extraction for urinary crystal detection. *Pattern Recognit Lett.* 2018;105:112-121.
9. Zhang Y, Zhang L, Qin L, Qin Z, Guo Y, Liu H, et al. Small dataset challenges in automated microscopic diagnosis. *J Med Syst.* 2020;44(8):143.
10. Li Y, Zhang L, Qin L, Qin Z, Guo Y, Liu H, et al. Automated throughput enhancement in clinical laboratories via machine learning. *J Clin Pathol.* 2020;73(5):260-268.
11. Chen H, Jiao X, Jin L, Zhao J, Tang Y. Limitations of handcrafted features in urine sediment morphology. *Biomed Signal Process Control.* 2020;58:101845.



12. Garcia R, Lopez M. High-precision classification of uric acid crystals using Inception-V3. *IEEE Trans Med Imaging*. 2022;41(8):2105-2118.
13. Kumar P, Rawat SS, Jain S. Application of ResNet architectures in renal sediment analysis. *Comput Biol Med*. 2021;135:104612.
14. Tan L, Wang J, Song Z, Ye C, Feng L, Zhang Y. Ensemble learning for the detection of rare pathological crystals. *Artif Intell Med*. 2022;124:102231.
15. Zhao X, Li L, Jiao X, Jin L, Zhao J, Tang Y. Attention-based neural networks for discriminating benign and pathological urine particles. *Med Image Anal*. 2023;85:102741.
16. Roboflow. Urine Crystals Dataset [Internet]. San Francisco (CA): Roboflow Universe; 2024 [cited 2026 Mar 10]. Available from: <https://universe.roboflow.com>
17. Long J, Wang J, Song Z, Ye C, Feng L, Zhang Y. Optimization of image preprocessing in automated microscopic analysis. *J Clin Lab Anal*. 2023;37(4):e24810.
18. Wong TT, Yeh PY. Reliable accuracy estimates of deep learning models in medical imaging with small datasets. *Inf Sci*. 2020;508:13-36.
19. Ramos P, Silva A. Deploying deep learning models for real-time pathology using FastAPI and Docker. *SoftwareX*. 2024;25:101612.
20. Liang Y, Kong G, Pu J, He B, Jin L, Xu J, et al. Automated identification of urinary particles using deep convolutional neural networks. *Comput Biol Med*. 2022;142:105198.
21. Wang L, Zhang L, Qin L, Qin Z, Guo Y, Liu H, et al. Performance of ResNet architectures in automated urinalysis. *IEEE Access*. 2021;9:112340-112352.
22. Aris M. Challenges in cross-platform generalization of automated urinalysis systems. *J Med Syst*. 2026;50(2):45-58.
23. Brown L. Robustness of AI models in contaminated clinical urine samples. *Clin Chem Lab Med*. 2026;64(1):112-120.
24. Choi S, Park J, Kim H, Lee J, Jung S, Moon T. Generalization gaps in deep learning models for urinary sediment analysis. *Digit Med Rev*. 2025;11(3):201-215.