ASEAN
Journal of Scientific and Technological Reports
Online ISSN:2773-8752

Research article

Comparison of Color Space Performance for Colorimetric Detection of Heavy Metals in Drinking Water Using Image Processing and Convolutional Neural Networks

Atthapol Phongsuwan[1], Nadh Ditcharoen[1*], and Sompong Valuvanathorn[1]

[1] Faculty of Science, Ubon Ratchathani University, Ubon Ratchathani, 34190, Thailand
* Correspondence: nadh.d@ubu.ac.th

Citation:
Phongsuwan, A.; Ditcharoen, N.; Valuvanathorn, S. Comparison of color space performance for colorimetric detection of heavy metals in drinking water using image processing and convolutional neural networks. *ASEAN J. Sci. Tech. Report.* **2026**, *29*(4), e260866. https://doi.org/10.55164/ajstr.v29i4.260866

Article history:
Received: August 16, 2025
Revised: February 11, 2026
Accepted: February 22, 2026
Available online: March 24, 2026

Abstract: Heavy metal contamination in drinking water, including arsenic, lead, cadmium, iron, copper, and manganese, is an environmental issue that adversely affects human health and ecosystems. Although standard laboratory methods provide high accuracy, they are limited by cost, time, and operational complexity. This study aimed to develop and compare the performance of eight color spaces (RGB, RGBA, Grayscale, GrayscaleA, CIE LAB, CIE LABA, HSV, HSVA) for classifying heavy metal concentration levels from microplate images using image processing techniques combined with a Convolutional Neural Network (CNN), and to select the optimal color space using the Weighted Sum Model. The dataset comprised images of six types of heavy metals, with 368,000 images after preprocessing and data augmentation. The model was trained using K-Fold Cross Validation (K=5). Experimental results showed that the HSVA color space achieved the best performance (Accuracy 99.65%, Loss 0.01382, Training Time 1,606 seconds). When tested on a separate set of 920 images, the model maintained an accuracy of 99.0%, indicating stability and strong practical applicability. The findings confirm that selecting an appropriate color space in conjunction with a CNN significantly improves the accuracy of heavy metal analysis on test datasets and has strong potential for further development into a portable, low-cost analytical tool for field applications.

Keywords: Heavy metals; image processing; color spaces; nvolutional neural network; colorimetric measurement

1. Introduction

Heavy metal contamination in drinking water—such as arsenic (As), lead (Pb), cadmium (Cd), iron (Fe), copper (Cu), and manganese (Mn)—poses a serious environmental threat with profound impacts on human health and ecosystems. Bioaccumulation of these metals can lead to neurological disorders, kidney dysfunction, cardiovascular diseases, and gastrointestinal complications [1, 2]. Current detection of heavy metals predominantly relies on laboratory-based standard methods, such as Atomic Absorption Spectrometry (AAS) and Inductively Coupled Plasma Mass Spectrometry (ICP-MS), which, although highly accurate, are limited by high operational costs, bulky instrumentation, and complex procedures, making them unsuitable for field deployment [3, 4]. To address these limitations, portable, low-cost colorimetric techniques have been explored, often using smartphones or digital cameras to capture color changes resulting from chemical reactions with heavy metals. Such methods

offer low cost, portability, and practical applicability in on-site testing [5, 6]. However, the accuracy of this approach depends critically on image quality and the choice of a suitable color space for chemical classification.

Previous studies have investigated the influence of color space transformations on the performance of Convolutional Neural Networks (CNNs) across various domains, including plant disease diagnosis, smoke detection, and medical image analysis [7 – 15]. Findings indicate that certain color spaces—such as HSV and CIE LAB—can significantly improve classification accuracy. Although prior studies have demonstrated that color space transformations can improve CNN performance across various domains, there remains a lack of a systematic investigation focused specifically on colorimetric heavy metal detection in drinking water. In particular, no prior study has comprehensively compared multiple conventional and alpha-channel color spaces within a single experimental framework. This gap limits the development of robust, field-deployable colorimetric detection systems.

Therefore, this study aims to compare the performance of eight color spaces—RGB, RGBA, Grayscale, GrayscaleA, CIE LAB, CIE LABA, HSV, and HSVA—in classifying heavy metal concentration levels from microplate images using image processing techniques combined with CNNs. The Weighted Sum Model (WSM) [16, 17] is employed to identify the optimal color space that achieves high classification accuracy while balancing processing time and loss. The results are intended to serve as a guideline for developing low-cost, portable heavy metal analysis tools for effective field applications.

2. Research Objectives

2.1 To develop a convolutional neural network model for classifying and analyzing heavy metal concentration levels in drinking water from microplate images.

2.2 To compare the performance of eight color spaces — RGB, RGBA, Grayscale, GrayscaleA, CIE LAB, CIE LABA, HSV, and HSVA — using accuracy, loss, and training time as evaluation metrics.

2.3 To identify the optimal color space using the weighted sum model and evaluate the model's performance on a previously unseen test dataset.

3. Materials and Methods

This research used Python 3.8, OpenCV, TensorFlow, and Keras to develop and evaluate CNN models. The methodology consisted of 4 steps, illustrated in Figure 1 and detailed as follows.

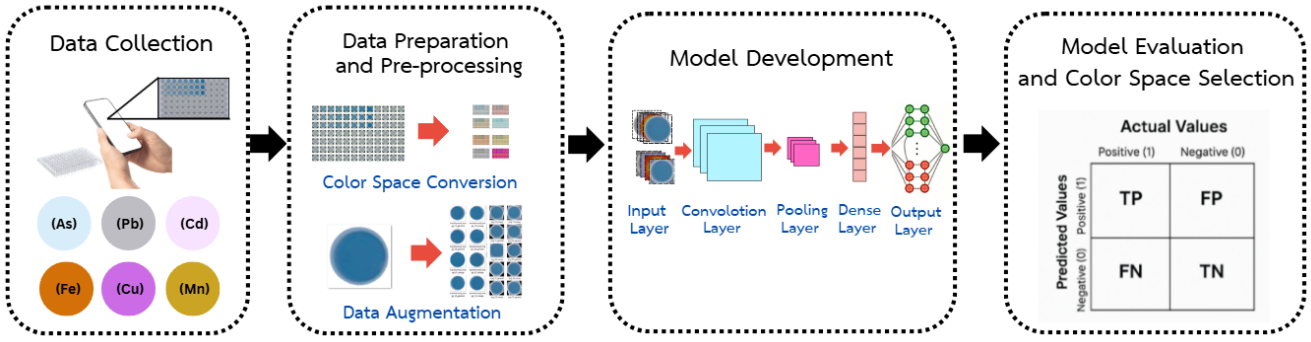


Figure 1. Research methodology.

3.1 Data Collection

The image data used in this study were obtained from Chunta [18], originating from laboratory analyses using a simple heavy metal test kit applied to drinking water samples. Tests were conducted on six types of heavy metals across 46 concentration levels, as shown in Table 1. Images were captured from a 96-well microplate using a Samsung SM-A515F smartphone with a resolution of 3,000 × 4,000 pixels, a 5 mm focal length lens (equivalent to 25 mm in 35 mm format), an aperture of f/2, a shutter speed of 1/365 s, and ISO-32. No flash was used, and the center-weighted average metering mode was applied. The camera was positioned

at 0° directly above the microplate, with constant lighting and fixed camera settings to ensure high-quality, consistent images, as illustrated in Figure 2.

Table 1. Details of the dataset [18].

Heavy Metal	Detection Range (mg/L)	# of Concentration Level (Class)
Arsenic (As)	0.007 – 4.00	10
Cadmium (Cd)	0.0002 – 0.05	9
Copper (Cu)	0.35 – 6.00	6
Iron (Fe)	0.10 – 100.00	9
Lead (Pb)	0.007 – 0.10	5
Manganese (Mn)	0.006 – 0.10	7

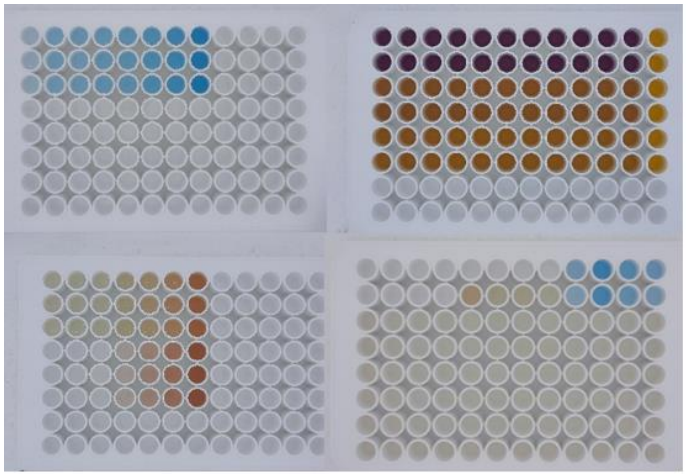
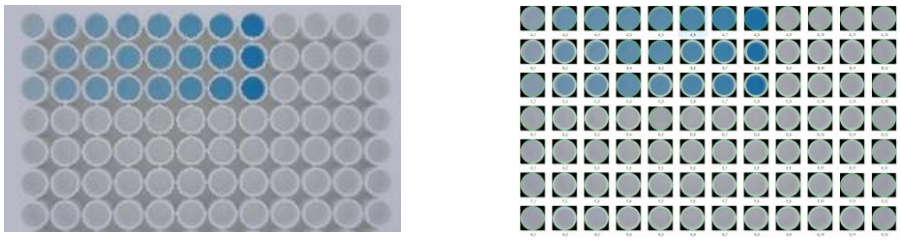


Figure 2. Example of heavy metal testing images using a 96-well microplate test kit [18].

3.2 Data Preparation and Pre-processing

3.2.1 Image Segmentation

The collected images were segmented by detecting and cropping the boundaries of the microplate wells using the Hough Circle Transform function (https://docs.opencv.org) [5]. The cropped images were then resized to 100×100 pixels. To enhance the CNN's learning capability, an alpha channel was created to make the area outside the circle transparent, and the images were saved as PNG files for further processing. Figure 3 presents examples of the images before and after pre-processing.



(A) (B)

Figure 3. Example of the images (A) before image segmentation and (B) after image segmentation.

3.2.2 Color Space Conversion

Each cropped image was converted into eight color spaces (RGB, RGBA, Grayscale, GrayscaleA, CIE LAB, CIE LABA, HSV, HSVA), as illustrated in Figure 4. Before being input into the CNN model, the pixel values were normalized to the range 0–1 to reduce data variability.

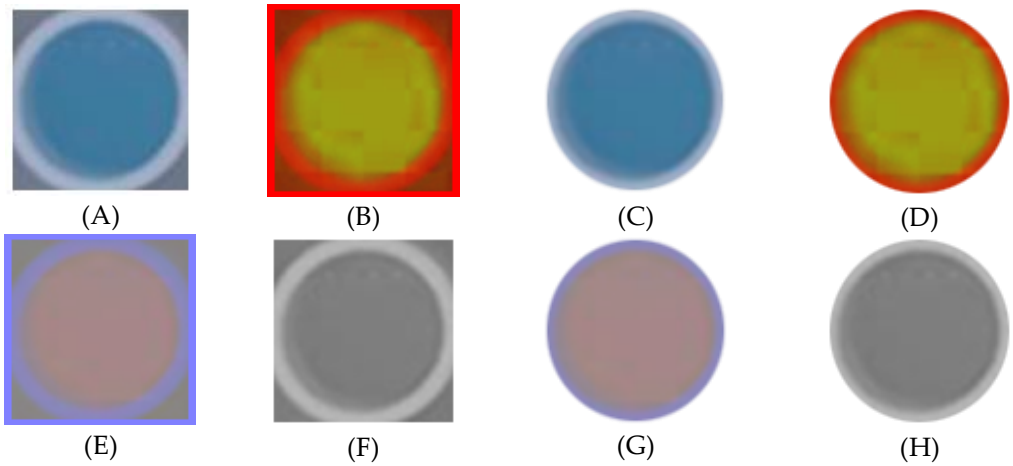


Figure 4. Example of the images after color space conversion (A) RGB, (B) HSV, (C) RGBA, (D) HSVA, (E) CIE LAB, (F) Grayscale, (G) CIE LABA, (H) GrayscaleA.

3.2.3 Data Augmentation

Data augmentation was applied to increase the size and diversity of the training dataset [19] by generating new images from the original images at a 9:1 ratio per class. The augmentation techniques included rotation, flipping, noise addition, brightness and contrast adjustment, and mix-up augmentation, as illustrated in Figure 5.

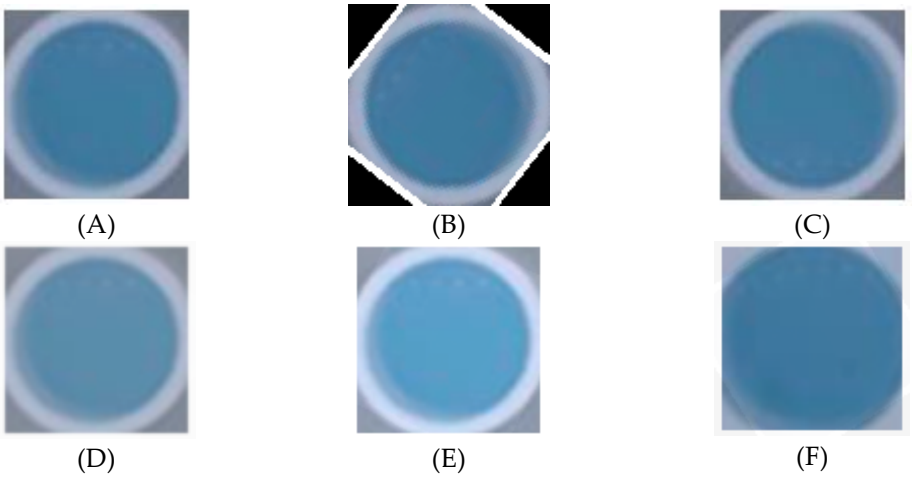


Figure 5. Example of the images after data augmentation (A) original image, (B) rotated image, (C) flipped image, (D) noise added image, (E) brightness and contrast adjustment image, (F) mix-up image.

3.2.4 Dataset Splitting

Each original image was augmented to produce nine additional images, resulting in 1,000 images per class (concentration level) when combined with the original. This yielded 46,000 images for each color space. For each color space dataset, the images were randomly split at an 80:20 ratio into a training set of 36,800 images and a validation set of 9,200 images. In this study, experiments were conducted across eight color spaces, yielding a total of 368,000 images used for training and validation.

3.3 Model Development

The CNN model was designed to support all 8 color spaces and comprised convolutional layers, max pooling layers, fully connected layers, and a dropout layer (rate = 0.5) before the final dense output layer with 46 nodes. The model used categorical cross entropy as the loss function, the Adam optimizer (learning rate = 0.001), and a batch size of 32. Training was conducted using K-fold cross-validation (K = 5) with early stopping (patience = 3). Details are provided in Table 2 and illustrated in Figure 6.

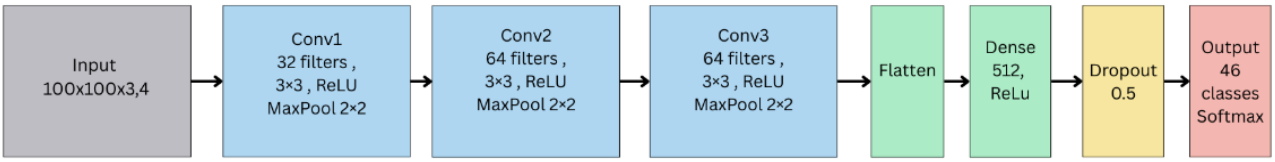


Figure 6. Structure and parameters of the convolutional neural network.

Table 2. Details of the convolutional neural network model.

Layer (Type)	Output	# of Parameters
Input	(100, 100, 3)	-
Conv2D (filters=32, kernel_size=(3,3), padding='same',activation='relu')	(98, 98, 32)	896
MaxPooling2D (pool_size=(2,2))	(49, 49, 32)	0
Conv2D (filters=64, kernel_size=(3,3), padding='same', activation='relu')	(47, 47, 64)	18,496
MaxPooling2D (pool_size=(2,2))	(23, 23, 64)	0
Conv2D (filters=128, kernel_size=(3,3), padding='same', activation='relu')	(21, 21, 128)	73,856
MaxPooling2D (pool_size=(2,2))	(10, 10, 128)	0
Flatten	12,800	0
Dense (activation='relu')	512	6,554,112
Dropout (rate=0.5)	512	0
Dense (activation='softmax')	46	23,598

3.4 Model Evaluation and Color Space Selection

The evaluation of color space performance and the selection of the optimal color space in this study comprised the following steps:

3.4.1 Train and test a separate CNN model for each of the eight color spaces using K-Fold Cross Validation (K = 5), recording key metrics including validation accuracy, validation loss, and training time for each color space.

3.4.2 Normalize the recorded metrics using Min–Max normalization to scale all values to the range of 0–1, enabling direct comparison.

3.4.3 Calculate the weighted sum score (Weighted Sum Model: WSM), as defined in Equation (1), by assigning weights to the metrics—accuracy = 0.5, loss = 0.3, and training time = 0.2—multiplying each normalized value by its corresponding weight, and summing the results to obtain a total score for each color space.

$$S_i = \sum_{j=1}^{n} w_j \times x_{ij} \quad (1)$$

where S_i is the weighted sum score for the i-th alternative, w_j is the weight assigned to the j-th criterion, x_{ij} is the normalized value of the i-th alternative for the j-th criterion, and n is the total number of criteria.

The weighting scheme was designed to reflect practical deployment priorities. Classification accuracy was assigned the highest weight (0.5) due to its direct impact on detection reliability. Validation loss (0.3) reflects model stability and convergence behavior, while training time (0.2) represents computational efficiency, which is important for future scalability but secondary to predictive performance.

3.4.4 Select the optimal color space by ranking all color spaces based on their WSM scores and choosing the one with the highest score.

3.4.5 Test the model trained on the selected color space using a testing set that was not included in the training process. Evaluate the results using a confusion matrix, which includes true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN), along with the metrics accuracy, precision, recall, and F1-score, as illustrated in Figure 7 and defined in Equation (2) - (5) [19], to confirm stability and practical applicability, as defined in the following equations.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

$$F1\text{-}score = \frac{2\times Precision\times Recall}{Precision+Recall} \tag{5}$$

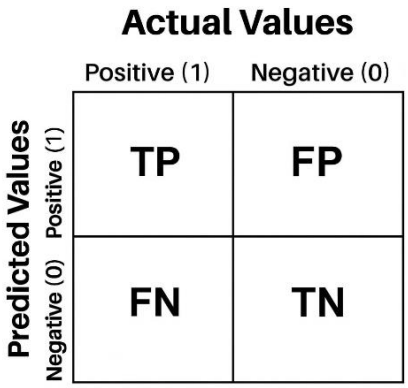


Figure 7. Confusion matrix and equations for evaluating model performance.

4. Results

4.1 Model Performance Evaluation

The CNN models were trained on eight different color spaces, and the HSVA color space achieved the best results, with the highest average accuracy (99.65%), the lowest loss (0.01382), and a training time of 1,606 seconds. This demonstrates a balance between classification accuracy and computational efficiency. The HSV and CIE LAB color spaces ranked second and third, respectively, while GrayscaleA recorded the lowest performance across both accuracy and loss. Results of average accuracy, loss, and training time for each color space are provided in Table 3. The low standard deviation across folds indicates stable model performance and minimal sensitivity to data partitioning.

Table 3. Results of average accuracy, loss, and training time for each color space.

Color Space	Validation Accuracy	Validation Loss	Training Time (s)
HSVA	**99.65 ± 0.25%**	**0.01382**	1606
HSV	99.32 ± 0.67%	0.02740	1318
CIE LABA	98.71 ± 1.11%	0.05510	1822
CIE LAB	99.43 ± 0.25%	0.01992	2090
RGBA	98.61 ± 1.24%	0.02554	1661
RGB	97.62 ± 1.10%	0.04800	**1176**
Grayscale	98.56 ± 0.89%	0.07150	1840
GrayscaleA	98.16 ± 1.40%	0.05210	2039

4.2 Weighted Sum Model (WSM) Analysis

By combining the normalized scores of accuracy, loss, and training time, HSVA achieved the highest weighted score (0.8819), followed by HSV (0.8662) and CIE LABA (0.8309). This confirms HSVA's balanced performance across all evaluation metrics. The normalized metrics and WSM analysis for each color space are provided in Table 4.

Table 4. Results of normalized metrics for each color space.

Color Space	Normalized Accuracy	Normalized Loss	Normalized Time	WSM
HSVA	1	0.8067	0.2316	**0.8819**
HSV	0.9967	0.6168	0.3694	0.8662
CIE LABA	0.9906	0.2294	0.1282	0.8309
CIE LAB	0.9978	0.7214	0	0.7430
RGBA	0.9896	0.6428	0.2053	0.7434
RGB	0.9796	0.3287	0.4373	0.7452
Grayscale	0.9891	0	0.1196	0.6155
GrayscaleA	0.985	0.2713	0.0244	0.4959

4.3 Performance Evaluation of the HSVA-based CNN Model on the Test Image Dataset

The HSVA-based CNN model was tested on 920 images (20 per class) that were not used during training. The model achieved accuracy, precision, recall, and F1-score values of 0.99. Errors occurred only minimally in classes with similar colors, as shown in the confusion matrix in Figure 8. Misclassifications mainly occur between adjacent concentration levels with similar color intensity.

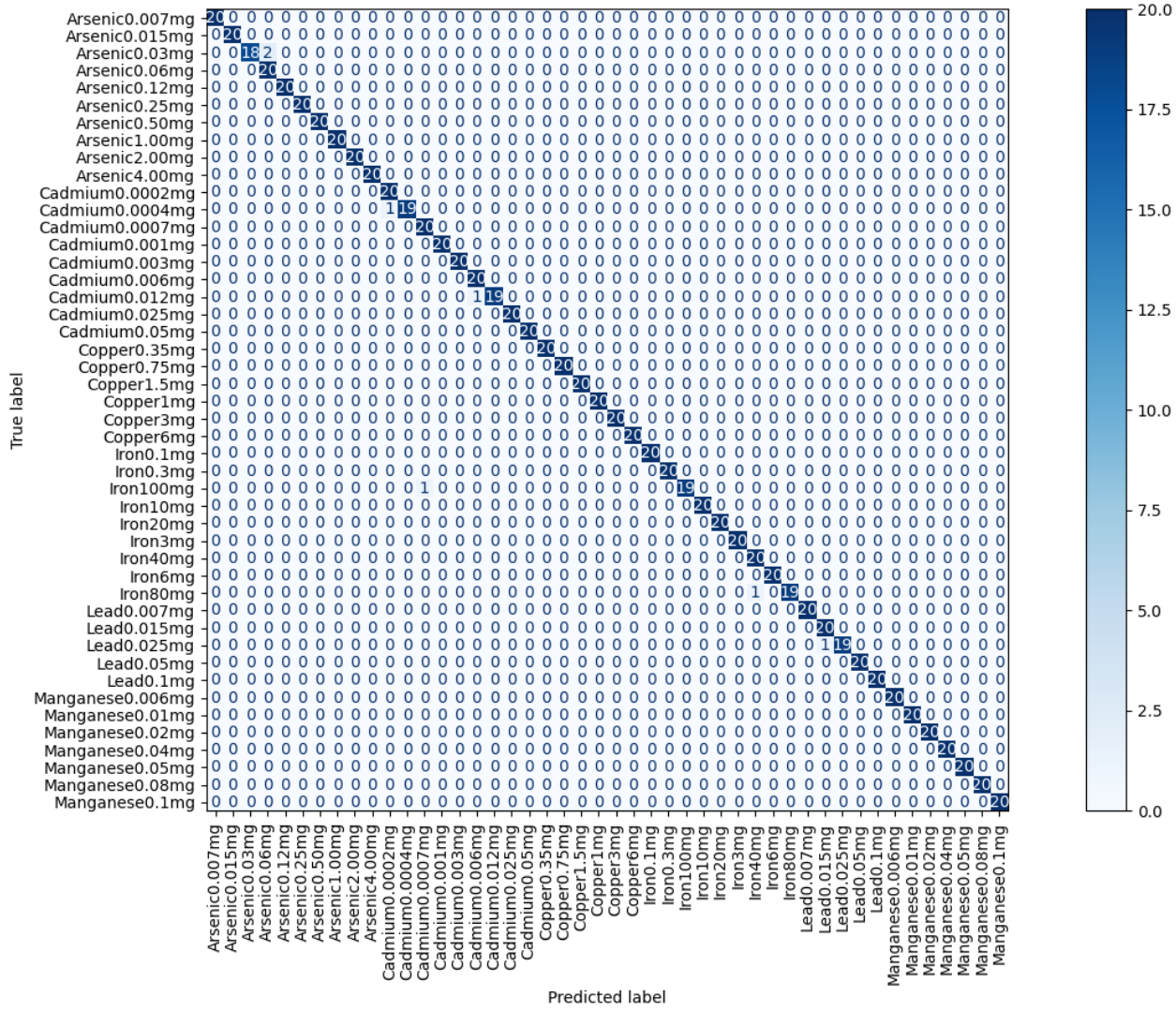


Figure 8. Confusion matrix of the HSVA-based CNN model on the test image dataset.

4.4 Key Findings

HSVA delivered the highest classification accuracy and demonstrated a balanced trade-off between accuracy, loss, and training time. Real-world testing confirmed the model's applicability and robustness in practical scenarios. Including the alpha channel in HSVA enabled the model to focus solely on the chemical reaction area, reducing background noise and improving classification accuracy.

5. Discussion

The results indicate that the HSVA color space achieved the highest performance in terms of accuracy, loss, and training time compared to other color spaces. Preserving the alpha channel enabled the model to focus solely on the chemical reaction area, reducing background noise and improving classification accuracy. This finding is consistent with Taithong *et al.* [5] and Xian *et al*. [12], who reported that using color spaces with fine color differentiation, combined with region-of-interest (ROI) masking, significantly improves CNN performance. Moreover, the application of data augmentation techniques—such as rotation, flipping, brightness and contrast adjustment, and mix-up augmentation—played a crucial role in reducing overfitting and improving the model's ability to process diverse real-world images. This observation aligns with Haikal *et al*. [9] and Liu [10], who emphasized that increasing the diversity of training data enhances model generalization. However, this study has certain limitations. The image dataset was collected under fixed lighting and camera angles, which may not fully represent all real-world scenarios. Expanding the experiments to include varying lighting conditions and image-acquisition settings could further improve the model's robustness and applicability in practical environments.

6. Conclusions

This study successfully achieved its objectives: developing a CNN model for classifying and analyzing heavy metal concentrations from microplate images; comparing eight color spaces; and selecting the optimal color space using the WSM approach. The HSVA color space was identified as the best performer, offering a balanced trade-off among accuracy, loss, and training time. Testing with real-world images confirmed the model's stability and practical potential. The proposed HSVA-based CNN framework can be directly extended to smartphone-assisted water quality monitoring, on-site screening in rural communities, and rapid preliminary assessment before laboratory confirmation. With further validation under varying lighting conditions, the method has potential for integration into portable diagnostic platforms and decision-support systems for environmental health monitoring.

7. Acknowledgements

The authors would like to express their gratitude to Assoc. Prof. Dr. Suthicha Chunta for kindly providing the dataset used in this research, and to Assoc. Prof. Dr. Purim Jarujamrus for valuable suggestions and guidance throughout the course of this study.

Author Contributions: Conceptualization, N.D. and S.V.; methodology, N.D., A.P., and S.V.; software, A.P.; validation, N.D., A.P., and S.V.; formal analysis, N.D. and A.P.; investigation, N.D.; resources, N.D. and A.P.; data curation, A.P.; writing—original draft preparation, N.D.; writing—review and editing, N.D., A.P., and S.V.; visualization, A.P.; supervision, N.D. and S.V.; project administration, N.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] Yang, Y.; Li, Y.; Wang, X.; Li, J.; Chen, W. The Gut Microbiota and Chronic Inflammation Induced by Heavy Metals: A Review. *Ecotoxicol. Environ. Saf.* **2023**, *254*, 114720. https://doi.org/10.1016/j.ecoenv.2023.114720

[2] Jaishankar, M.; Tseten, T.; Anbalagan, N.; Mathew, B. B.; Beeregowda, K. N. Toxicity, Mechanism and Health Effects of Some Heavy Metals. *Interdiscip. Toxicol.* **2014**, *7*(2), 60–72. https://doi.org/10.2478/intox-2014-0009

[3] Bansod, B. K.; Kumar, T.; Thakur, R.; Rana, S.; Singh, I. A Review on Various Electrochemical Techniques for Heavy Metal Ions Detection with Different Sensing Platforms. *Microchem. J.* **2017**, *133*, 502–512. https://doi.org/10.1016/j.microc.2017.03.025

[4] Fakayode, S. O.; Walgama, C.; Fernand Narcisse, V. E.; Grant, C. Electrochemical and Colorimetric Nanosensors for Detection of Heavy Metal Ions: A Review. *Sensors* **2023**, *23*(22), 9080. https://doi.org/10.3390/s23229080

[5] Taithong, P.; Wichai, S.; Waranusast, R.; Riyamongkol, P. Image-Based Automatic Counting of *Bacillus cereus* Colonies Using Smartphone. *Int. J. Adv. Comput. Sci. Appl.* **2022**, *13*(2), 627–634. http://dx.doi.org/10.14569/IJACSA.2022.0130274

[6] Chunta, S.; Phongthai, S.; Jarujamrus, P. Simple Colorimetric Assay Using Pectin Hydrogel Reagent Coupled with Camera-Based Photometry for Trace Arsenic Determination. *Anal. Bioanal. Chem.* **2023**, *415*, 4603–4614. https://doi.org/10.1007/s00216-023-04762-z

[7] Ahmad, N.; Singh, S.; AlAjmi, M. F.; Hussain, A.; Raza, K. CropGCNN: Color Space-Based Crop Disease Classification Using Group Convolutional Neural Network. *PeerJ Comput. Sci.* **2024**, *10*, e2136. https://doi.org/10.7717/peerj-cs.2136

[8] Aydın Atasoy, N.; Kura, İ. Classification of X-ray and CT Images in Different Color Spaces Using Robust CNN. *J. Eng. Sci. Des.* **2024**, *12*(3), 505–516. https://doi.org/10.21923/jesd.1415150

[9] Haikal, A. L. A.; Yudistira, N.; Ridok, A. Comprehensive Mixed-Based Data Augmentation for Detection of Rice Leaf Disease in the Wild. *SSRN* **2024**. https://doi.org/10.2139/ssrn.4705403

[10] Liu, X. *Evaluating the Impact of Data Augmentation on Explainable AI in Medical Image Analysis*; Master's Thesis, Eindhoven University of Technology, Eindhoven, The Netherlands, 2024.

[11] Taha, M.; Safaa, M. Impact of Colour Space Transformation on Smoke Detection Accuracy Using ResNet50. *Infotech Spectrum: Iraqi J. Data Sci.* **2025**, *2*(1), 37–49.

[12] Xian, Z.; Huang, R.; Towey, D.; Yue, C. Convolutional Neural Network Image Classification Based on Different Color Spaces. *Tsinghua Sci. Technol.* **2025**, *30*(1), 402–417. https://doi.org/10.26599/TST.2024.9010001

[13] Satrya, G. B.; Ramatryana, I. N. A.; Shin, S. Y. Compressive Sensing of Medical Images Based on HSV Color Space. *Sensors* **2023**, *23*, 2616. https://doi.org/10.3390/s23052616.

[14] Frincu, R. M. Artificial Intelligence in Water Quality Monitoring: A Review of Water Quality Assessment Applications. *Water Qual. Res. J.* **2025**, *60*(1), 164–176. https://doi.org/10.2166/wqrj.2024.049

[15] Olawade, D. B.; Wada, O. Z.; Ige, A. O.; Egbewole, B. I.; Olojo, A.; Oladapo, B. I. Artificial Intelligence in Environmental Monitoring: Advancements, Challenges, and Future Directions. *Hyg. Environ. Health Adv.* **2024**, *12*, 100114. https://doi.org/10.1016/j.heha.2024.100114

[16] Budiharjo; Windarto, A. P.; Muhammad, A. Comparison of Weighted Sum Model and Multi Attribute Decision Making Weighted Product Methods in Selecting the Best Elementary School in Indonesia. *Int. J. Softw. Eng. Its Appl.* **2017**, *11*(4), 69–90. http://dx.doi.org/10.14257/ijseia.2017.11.4.06.

[17] Ouchra, H.; Belangour, A. Object Detection Approaches in Images: A Weighted Scoring Model Based Comparative Study. *Int. J. Adv. Comput. Sci. Appl.* **2021**, *12*(8), 268–275. http://dx.doi.org/10.14569/IJACSA.2021.0120831

[18] Chunta, S. *Development of the Hydrogel-Based Reagent Field Test Kit for Screening Heavy Metal in Water Coupled with Smartphone-Based Colorimetric Analyzer*; Research Report, Agricultural Research Development Agency (Public Organization): Bangkok, Thailand, 2021.

[19] Nasritha, K.; Kerdprasop, K.; Kerdprasop, N. Comparison of Sampling Techniques for Imbalanced Data Classification. *J. Appl. Inform. Technol.* **2018**, *1*(1), 20–37. https://doi.org/10.14456/jait.2018.2