Health care waste classification using deep learning aligned with Nepal's bin color guidelines

Suman Kunwar*[a] and Prabesh Rai[b]

[a]DWaste, Baltimore, MD, USA

[b]Department of Business & Information Technology, Lambton College of Applied Arts and Technology, Ottawa, ON, Canada

Abstract

The increasing number of Health Care facilities in Nepal has added up the challenges on managing health care waste (HCW). Improper segregation and disposal of HCW leads to contamination, spreading of infectious diseases and risk for waste handlers. This study benchmarks the state of the art waste classification models: ResNeXt-50, EfficientNet-B0, MobileNetV3-S, YOLOv8-n and YOLOv5-s using stratified 5-fold cross-validation technique on combined HCW data. YOLOv5-s achieved the highest accuracy (95.06%) but fell short with the YOLOv8-n model in inference speed with few milliseconds. The EfficientNet-B0 showed promising results of 93.22% accuracy but took the highest inference time. Following a repetitive ANOVA test to confirm the statistical significance, the best performing model (YOLOv5-s) was deployed to the web with bin color mapped using Nepal's HCW management standards. Further work is suggested to address data limitation and ensure localized context.

Keywords: Health care waste; Deep learning; Waste classification; Computer vision; Nepal health care waste standards.

1. Introduction

Managing HCW is a pressing challenge fueled by increasing waste generation by health care facilities. Improper segregation and disposal of HCW results in the spread of infectious diseases, contamination and puts risk towards waste handlers. Traditionally, HCW are done manually which is often labor-intensive, prone to human errors and poses a threat to waste handlers. Numerous researchers have demonstrated the feasibility of AI powered solutions for automated waste sorting and its management.

Studies have applied deep learning (DL) ResNeXt-50 [1] and EfficientNet_B7 [2] models, with transfer learning to get higher accuracy results in HCW classifications, even with small datasets. The intersection of IoT with models like YOLOv5 [3], YOLOv8 [4] and MobileNetV3 [5] has shown promising results on automated sorting of medical waste showcasing real world implementation.

In the study, Zhou et al. used ResNeXt-50 and achieved 97.2% accuracy on an eight-class private dataset, while other fine tune transfer learning on EfficientNet-B7, reaching an accuracy of 99% on six class dataset. Automated sorting with the help of sensors, robotic arms and edge devices has also shown promising results. These approaches have reported with 97.93% accuracy highlighting the real-world deployment in healthcare environments [6]. Moktar et al. used YOLO combined with IoT for the detection of HCW and achieved a mAP of 98% using 6,7860 images.

With the promising results, challenges like small dataset size, poor image quality and controlled environment testing persists raising concerns [7] on the scalability, infrastructure feasibility and adoption with existing waste management systems. These put forward the needs of expansion of dataset to generalize the detection, needs of real time deployment with real world settings and integration of AI with IoT for automation.

In Nepal, there are 16,611 health care facilities [8] serving people that follows to the guidelines set by National Health Care Waste Management Standards and Operating Procedures to manage HCW. These guidelines divide waste in General Health Care Waste (General HCW) and Hazardous Health Care Waste (Hazardous HCW). The Hazardous HCW spans in multiple categories and is mapped with various color codes as depicted in Fig. 1.

The mapping of individual items based on waste category, label and color codes are shown in Table 1.

A study by Patel et al. on HCWM practices and knowledge in 10 government district hospitals in Madhesh Pradhesh found that the waste handlers showed low adherence to HCWM guidelines and lacked sufficient knowledge, despite having adequate facilities for managing HCM [9]. This poses a risk towards waste handlers and residents nearby hospitals as highlighted by Karki et al. study [10]. Furthermore, some studies suggest that with proper sorting the amount of HCW can be reduced in Nepal's context [11]. An example of waste bins from one of the hospitals is shown in Fig. 2.

The studies demonstrate the effectiveness of DL models on HCW segregation, suggesting that proper sorting can help in waste minimization and de-risk waste handlers. This study, therefore, explores state of the art waste classification models: YOLOv8-n, MobileNetV3-S, ResNeXt-50, EfficientNet-B0 and YOLOv5-s using stratified K-fold cross-validation method, which is preferred for optimizing sample selection and robust prediction results [12]. The best performing model mapped with Nepal bin color guidelines is deployed for public usage.

2. Materials and methods

The section discusses the dataset used in this study and the methods used to benchmark the models.

---

*Corresponding author. Email: sumn2u@gmail.com

Figure 1: Color codes for waste segregation.

Table 1: Healthcare waste categories, types, color codes, and items.

Waste category	Type	Color code	Items
Non-risk HCW	Biodegradable	Green	Organic wastes; peels of veges and fruits; rotten or stale foods
	Non-Biodegradable	Blue	Paper (A4, duplex, medicine covers, normal papers); plastic (polythene bags, wrappers, packaging plastic materials, cling wraps); bottles(water, soft drinks); vials; saline bottles
Risk HCW	Infectious	Red	Cotton, gauze, used gloves, IV sets, blood bags, urine bags, used PPE, diapers, pads, sputum containers, collection tubes, dialyzer
	Pathological		Body organs or tissues; placenta
	Cytotoxic		Unused or expired cytotoxic drugs; chemotherapy agents
	Pharmaceutical		Expired medicines
	Sharps		Needles, blades, scalpels, broken glass, lancets, suture needles
	Chemical	Yellow	Discarded chemicals, reagents, solvents, disinfectants
	Radioactive	Black	X-ray films; radioisotopes

Figure 2: Waste bins at Bhaktapur Cancer Hospital, Nepal.

2.1. Dataset and preprocessing

We use a combined dataset: Medical Waste Dataset 4.0 [13] and Pharmaceutical and Biomedical Waste dataset [14] to cover major categories of Nepal HCW. Medical Waste Dataset 4.0 was acquired from Tuscany Region, using OAK 4.0 camera device and span across: gauze, glove pair (latex, nitrile, surgery), glove single (latex, nitrile, surgery), medical cap, medical glasses, shoe cover pair, shoe cover single, test tube and urine bag. The RGB images are of size 1920 x 1080. The Pharmaceutical and Biomedical Waste dataset is collected by the Engineering UBU and consists of: Body Tissue or Organ, Organic wastes, equipment-packaging (plastic, paper, metal, glass), Syringe needles, Gauze, Gloves, Mask, Syringe and Tweezers.

In the Pharmaceutical and Biomedical Waste dataset, the Gloves contain mixed gloves images, a more generalized class. As the sep-

arated types of gloves are already present in the Medical Waste Dataset 4.0, we removed this class to remove the bias. The dataset is divided into 5 stratified folds; each fold containing 80% train and 20% validation data, and can be found in Kaggle[1]. The over-sample class images were reduced using the median class count value to handle the synthetic bias. For under sampling classes, data augmentation techniques such as flipping and brightness contrast were used.
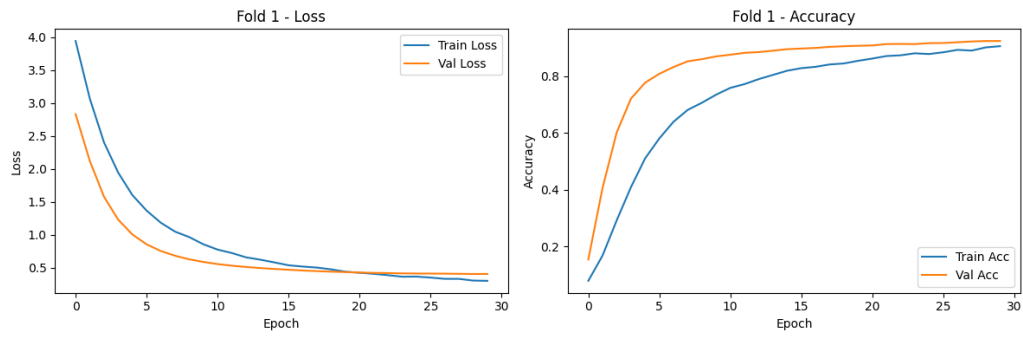
2.2. Model training

The experiments were conducted using two NVIDIA Tesla T4 GPUs. Models were tested using 5-fold stratified cross validation, with each fold trained for 30 epochs using MobileNetV3-S, ResNeXt-50, EfficientNet-B0 and YOLO models (v5-s/v8-n). The stratified K-fold technique was used to maintain the same percentage of instances for each label [15].

DL models such as MobileNetV3-S, ResNeXt-50 and EfficientNet-B0, used pre-trained weights trained in ImageNet dataset with frozen base model layers and a custom classification head. In contrast, YOLO models (v5-s/ v8-n) are trained from scratch. The precision, recall, F1-score and accuracy of each model along with inference time are reported as weighted average. Fig. 3. depicts the training and validation loss and accuracy for each fold of EfficientNet-B0 model.

The repeated measures of ANOVA results were also compared to determine statistical significance. When significance was found, Tukey's HSD Post-Hoc test was used to identify differing pairs while controlling the Type I error rate [16].
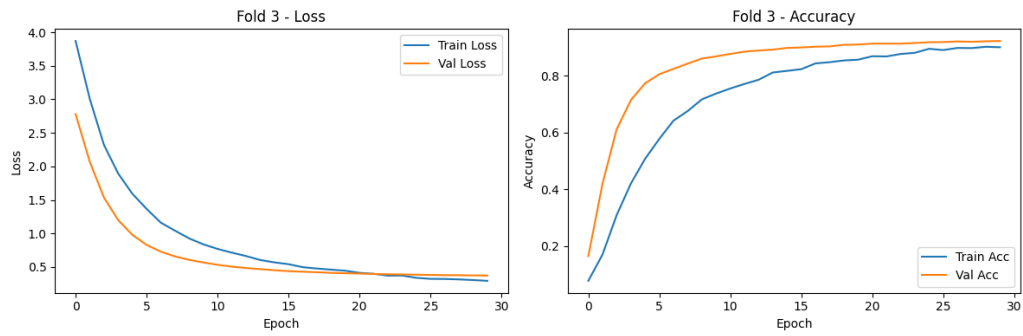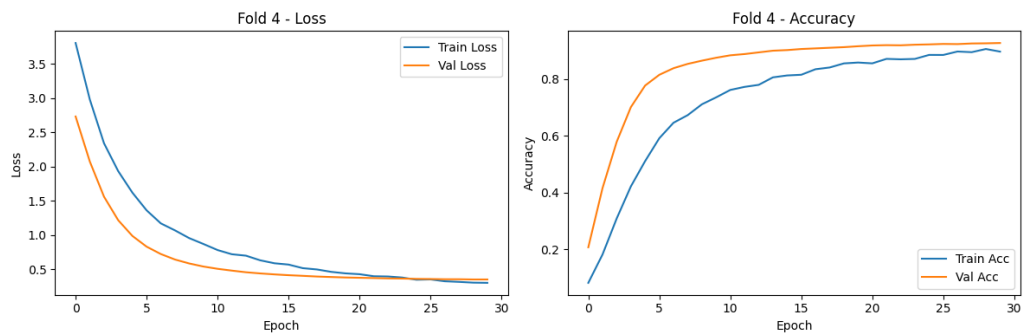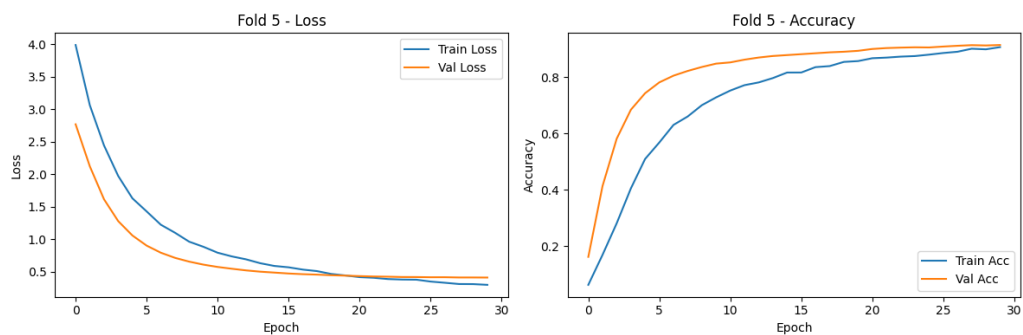
---

[1] https://www.kaggle.com/datasets/sumn2u/medical-waste

(a) Fold 1.
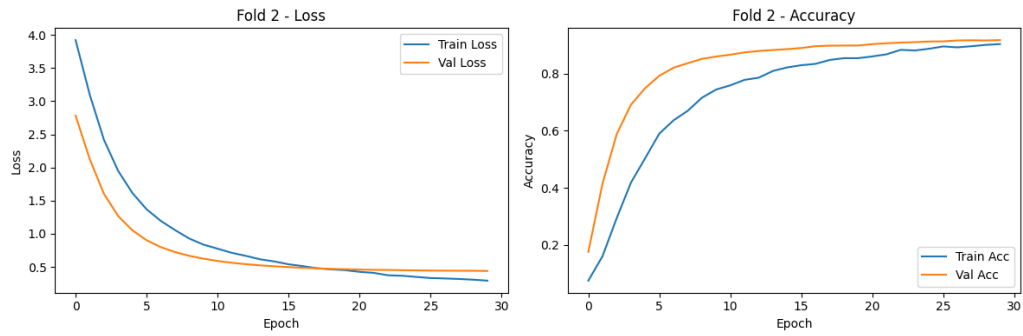
(b) Fold 2.

(c) Fold 3.

(d) Fold 4.

**(e)** Fold 5.

Figure 3: Training loss and accuracy curves for each fold in 5-fold cross-validation of EfficientNet-B0 model.

Model	Accuracy	Precision	Recall	F1-Score	Inference time (ms)
YOLOv8-n	0.9468	0.9644	0.9468	0.9457	**9.29**
MobileNetV3-S	0.9105	0.9290	0.9105	0.9095	369.24
ResNeXt-50	0.7451	0.7653	0.7451	0.7448	395.74
EfficientNet-B0	0.9322	0.9481	0.9322	0.9304	444.67
YOLOv5-s	**0.9506**	**0.9665**	**0.9506**	**0.9487**	10.97
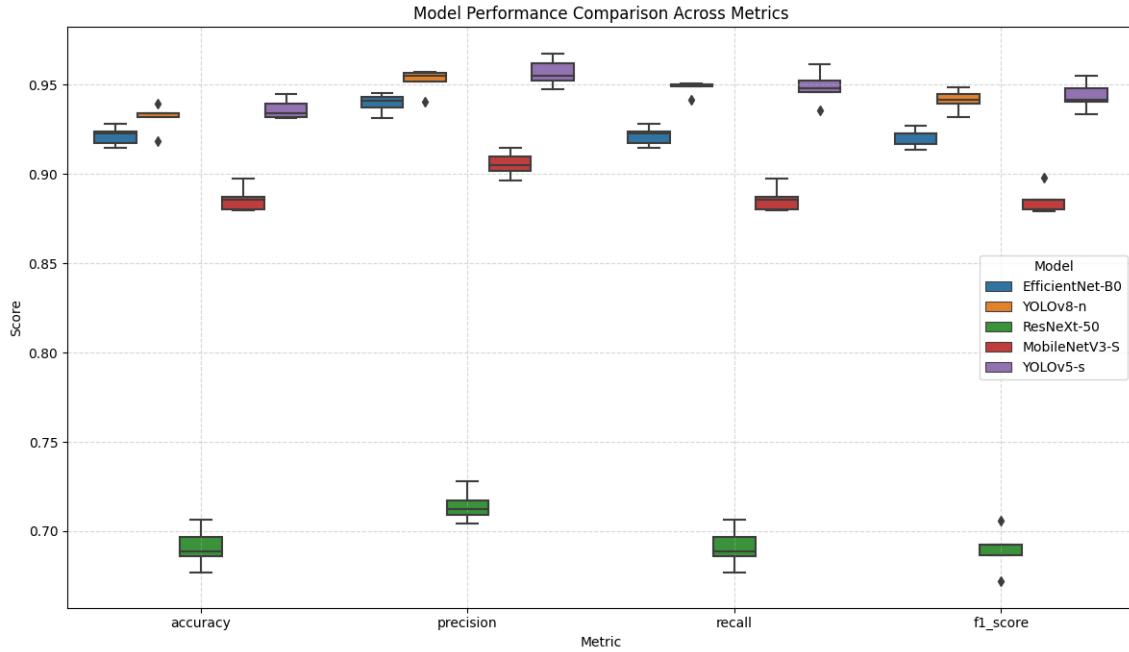
Figure 4: Model performance comparison across metrics using 5 fold.

3. Results and discussion

The experiment's results are shown in Table 2. YOLOv5-s, YOLOv8-n and EfficientNet-B0 outperformed MobileNetV3-S and ResNeXt-50 in most metrics. YOLOv5-s showed better performance than others but fell behind the YOLOv8-n model in inference time. YOLOv5-s achieved the highest accuracy of 95.06%; in comparison, the ResNeXt-50 achieved 74.51% accuracy making it the least accurate. MobileNetV3-s is quite competitive but fell short behind the top performers.

YOLOv8-n and YOLOv5-s showed lower inference time than other models and they achieved better results across accuracy, precision, recall, or F1-score are priorities followed by EfficientNet-B0. In real world applications where inference time is crucial, YOLO (v5-s/ v8-n) models are the best choice. The training time of EfficientNet-B0 is much higher compared to other models along with the model size.

The repeated ANOVA measures across the 5 stratified folds confirmed the highly significant effect of the model on the performance metrics. A significant difference between metrics themselves and the interaction between model and metrics was also highly significant. These results indicate that both the choice of model and type of evaluation metric can influence the performance, suggesting some models perform better on specific metrics than others. The model performance comparison across various models using different metrics can be seen in Fig. 4, which clearly shows that the YOLO models were outperforming the rest.

Our results were compared with the study that introduced the Medical Waste Dataset 4.0 shown in Table 3.

Table 3: Comparison of studies and their best-performing models.

Study	Number of classes	Best model	Accuracy
Bruno et al.	7 classes	EfficientNet-B0	99.45%
Our Work	23 classes	YOLOv5-s	95.06%

While the work by Bruno et al. achieved a high accuracy of 99.45% with EfficientNet-B0, its scope was limited to only seven classes. Our approach, utilizing the YOLOv5-s model, reached an accuracy of 95.06% on a significantly more challenging 23-class task, demonstrating robust and scalable performance. These results set a new standard for detailed medical waste sorting and confirm that YOLO models are useful for real-time sorting.

From our experiment, we found that YOLOv5-s performed well and is deployed in a Hugging Face space for real-world classification of waste images with bin mapping based on Nepal's guidelines[2]. Fig. 5 shows a sample HCW image being classified with bin color.

Our study has some limitations, including missing some classes: cytotoxic, radioactive, soiled and pathological, chemical and liquid waste that are included in Nepal's waste system. Additionally, the combined dataset is skewed towards common items such as gloves and gauze and is collected in non-Nepali settings. There is still a high chance that the model might fail in real world as waste often appears in occluded, mixed, and cluttered packaging. To mitigate these issues, further work needs to be done especially on data collection that represents generalized data in local settings.
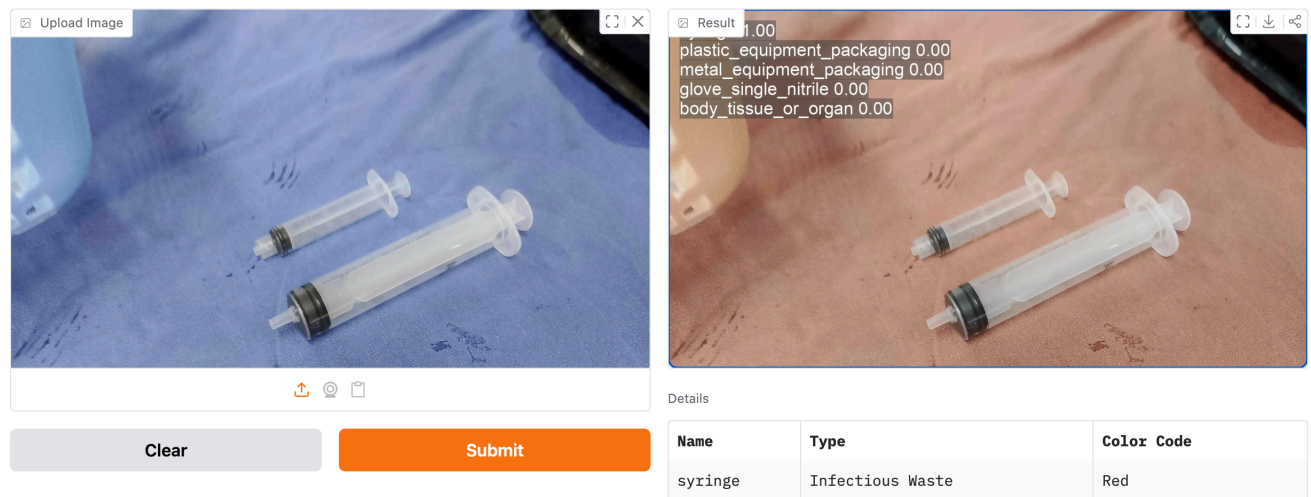
---

[2] https://huggingface.co/spaces/iamsuman/medical-waste-detector

Figure 5: HCW detection sample app with bin color based on Nepal's guidelines.

4. Conclusion

This study benchmarked state-of-the-art deep learning models (ResNeXt-50, EfficientNet-B0, MobileNetV3-S, YOLOv8-n, and YOLOv5-s) for automated HCW classification aligned with Nepal's national bin color guidelines, using a stratified 5-fold cross-validation approach on a combined HCW dataset. The YOLO models (YOLOv5-s and YOLOv8-n) demonstrated superior performance, with YOLOv5-s achieving the optimal balance between speed and accuracy (95.06% accuracy) and a low inference time (10.97 ms), making it highly suitable for real-time deployment. Although YOLOv8-n offered marginally faster inference (9.29 ms), and EfficientNet-B0 also performed competitively, both models were less efficient than YOLOv5-s when considering the trade-off with model size and overall inference time. Statistical validation via repeated ANOVA confirmed the significance of model selection on performance metrics.

The practical viability of this approach is demonstrated by a deployed web application integrated with Nepal's HCW color-coding system. While the current dataset lacks certain key Nepali waste categories (e.g., cytotoxic and chemical wastes) and localized context, this research offers a scalable, AI-driven framework for safer HCW management in Nepal. Future work should focus on expanding localized data collection, incorporating more waste categories, and integrating with IoT sorting mechanisms to further automate the system.

Acknowledgments

References

[1] Zhou H, Yu X, Alhaskawi A, Dong Y, Wang Z, Jin Q, Hu X, Liu Z, Kota V G, Abdulla M H A H, Ezzi S H A, Qi B, Li J, Wang B, Fang J & Lu H, A deep learning approach for medical waste classification, *Scientific Reports*, 12(1) (2022) 2159. ISSN 2045-2322. https://doi.org/10.1038/s41598-022-06146-2.

[2] Kumar N M, Mohammed M A, Abdulkareem K H, Damasevicius R, Mostafa S A, Maashi M S & Chopra S S, Artificial intelligence-based solution for sorting COVID related medical waste streams and supporting data-driven decisions for smart circular economy practice, *Process Safety and Environmental Protection*, 152 (2021) 482–494. ISSN 09575820. https://doi.org/10.1016/j.psep.2021.06.026.

[3] Lahoti J, Sn J, Krishna M V, Prasad M, Bs R, Mysore N & Nayak J S, Multi-class waste segregation using computer vision and robotic arm, *PeerJ Computer Science*, 10 (2024) e1957. ISSN 2376-5992. https://doi.org/10.7717/peerj-cs.1957.

[4] Moktar M H, Mohamed H, Hajjaj S S H & Baharuddin M Z, Medical waste sorting machine development with IoT and YOLO model utilization, *Journal of Engineering and Applied Science*, 72(1) (2025) 86. ISSN 1110-1903, 2536-9512. https://doi.org/10.1186/s44147-025-00661-5.

[5] Sarkar O, Dey A & Malik T, Modernized management of biomedical waste assisted with artificial intelligence, *International Journal of Biomedical and Clinical Analysis*, 3(2) (2023) 69–86. ISSN 25639218. https://doi.org/10.61797/ijbca.v3i2.265.

[6] Karmakar S K, Mallick M A, Middya A I & Roy S. AI powered medical waste management with IoT. In: *2024 IEEE Calcutta Conference (CALCON)*. IEEE, Kolkata, India (2024). ISBN 9798331516666, pp. 1–6. https://doi.org/10.1109/CALCON63337.2024.10914143.

[7] Bian X, Chen Y, Wang S, Cheng F & Cao H. Medical waste classification system based on OpenCV and SSD-mobilenet for 5G. In: *2021 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*. IEEE, Nanjing, China (2021). ISBN 9781728195070, pp. 1–6. https://doi.org/10.1109/WCNCW49093.2021.9420036.

[8] Ministry of Health and Population, Nepal. National health care waste standard operating procedure 2020. Government report (PDF) (2020). URL http://climate.mohp.gov.np/attachments/article/166/National%20Health%20Care%20Waste%20Standard%20Operating%20Procedure-2020.pdf, kathmandu, Nepal. Accessed: 2025-08-09.

[9] Patel J N, Upadhyay S K, Rajbhandari A, Bhandari R & Poudyal A, Knowledge about handling hazardous materials as factors associated with adherence to healthcare waste management practices among waste handlers at government district hospitals of Madhesh province, Nepal: A quantitative-qualitative methods study, *PLOS Global Public Health*, 4(12) (2024) e0002028.

ISSN 2767-3375. https://doi.org/10.1371/journal.pgph.0002028.

[10] Karki S, Niraula S R, Yadav D K, Chakravartty A & Karki S, Risk perception towards healthcare waste among community people in Kathmandu, Nepal, *PLOS ONE*, 15(3) (2020) e0230960. ISSN 1932-6203. https://doi.org/10.1371/journal.pone.0230960.

[11] Pathak D R, Nepal S, Thapa T, Dhakal N, Tiwari P & Sinha T K, Capacity assessment and implementation analysis of common treatment facility for the management of infectious healthcare waste in rapidly urbanising city of Nepal, *Waste Management & Research: The Journal for a Sustainable Circular Economy*, 39(1_suppl) (2021) 64–75. ISSN 0734-242X, 1096-3669. https://doi.org/10.1177/0734242X211013910.

[12] Pachouly J, Ahirrao S, Kotecha K, Selvachandran G & Abraham A, A systematic literature review on software defect prediction using artificial intelligence: Datasets, Data Validation Methods, Approaches, and Tools, *Engineering Applications of Artificial Intelligence*, 111 (2022) 104773. ISSN 09521976. https://doi.org/10.1016/j.engappai.2022.104773.

[13] Bruno A, Caudai C, Leone G, Martinelli M, Moroni D & Crotti F. Medical waste sorting: a computer vision approach for assisted primary sorting. In: *2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*. IEEE, Rhodes Island, Greece (2023). ISBN 9798350302615, pp. 1–5. https://doi.org/10.1109/ICASSPW59220.2023.10193520.

[14] Rapeepan Pitakaso, Surajet Khonjun, Thanatkij Srichok, Kanchana Sethanan, Natthapong Nanthasamroeng, Chawis Boonmee, Sarayut Gonwirat & Peerawat Luesak. Pharmaceutical and biomedical waste (2025). https://doi.org/10.34740/KAGGLE/DS/4201777.

[15] Kaliappan J, Bagepalli A R, Almal S, Mishra R, Hu Y C & Srinivasan K, Impact of cross-validation on machine learning models for early detection of intrauterine fetal demise, *Diagnostics*, 13(10) (2023) 1692. ISSN 2075-4418. https://doi.org/10.3390/diagnostics13101692.

[16] McHugh M L, Multiple comparison analysis testing in ANOVA, *Biochemia Medica* (2011) 203–209. ISSN 18467482. https://doi.org/10.11613/BM.2011.029.