

RESEARCH ARTICLE

Pill Recognition via Deep Learning Approaches

Mohd Rais Hakim Ramlee, Ismail Mohd Khairuddin*, Zubaidah Zamri, Muhammad Nur Aiman Shapiee and Muhammad Amirul Abdullah

Faculty of Manufacturing and Mechatronic Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, 26600 Pekan, Pahang, Malaysia.

ABSTRACT - Deep learning significantly transforms pill imaging recognition in the healthcare and pharmaceutical industries by automating the identification and classification processes based on visual indicators. It is important to develop a robust deep learning framework to ensure the accurate dispensing of medications. The features such as size, color, shape, markings and text imprint are scrutinized by these methods. However, real-world matching is difficult due to factors like the similarity of pill forms and the scarcity of databases. The goal of this work is to improve deep learning models for better classification of pill images. A dataset of 994 images are utilized from a public pharmaceutical database which sorted by 20 common type of pills. These images were split into training, validation, and testing sets in a 70:15:15 ratio. There are three different models which are YOLOv3, YOLOv5, and YOLOv8 were employed to the system. These models use performance metrics like recall, mean Average Precision (mAP), and precision as results. According to our results, YOLOv8 did remarkably well, obtaining a precision and F1-score of 99.17% and 96.95%, respectively, while YOLOv5 great with mAP and recall of 94.83% and 95%, respectively, outperforming the YOLOv3 model. The success of YOLOv8 underscores its significance in reducing medical errors with its accurate, real-time capabilities for identifying pills. The use of artificial intelligence in pill recognition not only lowers the chance of incorrect medication use but also streamlines the duties of healthcare professionals. This shift allows them to prioritize crucial responsibilities and simplifies the process of pill identification.

ARTICLE HISTORY

Received : 16th July 2024
Revised : 14th Sept 2024
Accepted : 20th Oct 2024
Published : 6th Nov 2024

KEYWORDS

Pill recognition
Deep learning
Pharmaceutical
Image processing
Pill imaging
YOLOv8

1.0 INTRODUCTION

Modern medicine relies heavily on tablet and capsule formulations as a cornerstone of drug administration. These small, precisely crafted dosage forms offer a reliable means of delivering therapeutic agents to patients. The design of pills varies widely, with each type tailored to specific medical needs, absorption rates, and ingredient compositions [1]. The period from the middle 1800s to the early 1900s saw a significant rise in pill usage, driven by technological advancements, regulatory changes, and the growth of pharmaceutical companies [2]. In this period, a large-scale manufacturing and broad adoption of pills are led by the significant technological progress, shifts in regulations, and the rise of pharmaceutical companies [3]. A significant progress has been made with the intersection of technology and healthcare, especially in the area of pill imaging recognition [4]. An increasingly important field as chronic diseases and complex medication regimens become more prevalent [5]. Traditionally, pill identification relied on manual inspection by healthcare professionals, which was time-consuming and error-prone [6]. Deep learning techniques have revolutionized pill recognition, enabling automated and efficient systems [7]. These advanced systems offer numerous advantages over conventional methods, including rapid analysis and identification, reduced verification time, and improved accuracy [8]. This enhanced precision contributes to increased patient safety. Moreover, these systems are adaptable, continuously improving their performance through learning from new data [9].

Accurate pill identification is crucial, especially for patients taking multiple medications with varying appearances [9]. This complexity increases the risk of mix-ups and dosage errors. Automated recognition systems help reduce these risks by providing consistent and reliable identification, ensuring patients take their prescribed medications correctly [10]. Furthermore, these systems play a vital role in maintaining product quality, identifying counterfeit drugs, and meeting regulatory requirements in the pharmaceutical industry [11]. Besides, they used in manufacturing to verify correct production and packaging, upholding quality control standards. In distribution, they help detect and prevent fake medication circulation, safeguarding patient health and supply chain integrity. These systems also assist in meeting the stringent verification processes required by regulatory bodies to ensure medication safety and effectiveness [12]. To address the challenges in pill image recognition, deep learning has emerged as a powerful approach. This method uses

artificial neural networks to identify complex patterns and features directly from data, offering a promising solution to this intricate task [13].

CNN, RNN, and other deep learning architectures have shown exceptional abilities in image recognition and sequential data analysis, making them ideal for the complexities of pill image recognition [14]. Deep learning models, especially CNNs, can process large amounts of visual data and learn to identify pills with high accuracy [15]. These models are trained on extensive datasets, allowing them to recognize a wide variety of medications, including those with subtle appearance differences. These deep learning techniques use large-scale datasets to learn complex patterns and features in pill images, enabling automated identification with high accuracy and efficiency [16]. The combination of deep learning with advanced imaging technologies, such as high-resolution cameras and mobile devices, has further improved the development and use of pill recognition systems in various healthcare settings [17]. This integration has led to more precise and efficient medication identification, which is essential for patient safety and improved healthcare delivery. As a result, healthcare professionals can use these advanced tools for quick and accurate medication verification, ultimately improving patient care and streamlining pharmaceutical management processes.

The advancement of pill image recognition, propelled by deep learning techniques, showcases how the merger of technology and healthcare can bring about significant improvements in patient care and safety. As these technologies continue to develop, they are expected to further refine the accuracy and efficiency of medication management. This progress will likely lead to better protection of patient health, improved treatment results, and stronger integrity in pharmaceutical production and distribution processes [18].

1.1 RELATED WORK

This research presents a detailed overview of developing an Attention-YOLO (AY) deep learning model specifically for recognizing round pill shapes. [19]. The study's introduction highlights the difficulty in distinguishing between similar round pills and emphasizes the need for automated recognition systems. In the methodology section, the researchers briefly explain how the AY model combines YOLOv3 (a popular object detection algorithm), an attention mechanism, and a hyper column architecture. For those interested in the technical specifics, the full paper provides more in-depth information about the model's structure. The study's findings are promising, with the AY model demonstrating excellent performance. It achieved a high accuracy rate of 92.28% when tested on the designated test dataset. This result suggests that the AY model could be a valuable tool for improving pill identification accuracy, particularly for round-shaped medications.

On the other research paper begins by presenting statistics on China's aging population and the increasing demand for medication, emphasizing the practical need for advanced pill identification technology [20]. The methodology section provides a clear outline of how the study was conducted, including dataset creation, image annotation process, model training procedures, and evaluation metrics used. To ensure real-world applicability, the researchers compiled a new dataset consisting of over 5,000 images, capturing a wide range of pills commonly encountered in practice. The study's results indicate that the YOLOv3 model achieves an optimal balance between accuracy and speed for real-time pill identification, with 80.69% Mean Average Precision (MAP) and 51 Frames Per Second (FPS). To put YOLOv3's performance in context, the researchers also compared it with two other object detection models which are RetinaNet and SSD, helping to understand YOLOv3's relative strengths and weaknesses in the context of pill identification.

This research introduces PGPNet, a novel framework designed for multi-pill detection in real-world images, addressing the complex challenge of simultaneously detecting and identifying multiple pills in a single image [21]. Unlike previous studies that focused on single pill classification, PGPNet's key innovation is its use of external knowledge, such as co-occurrence relationships and relative pill sizes, to differentiate between visually similar pills. The framework constructs three heterogeneous graphs - co-occurrence, relative size, and visual semantic - which are integrated with visual features to enhance detection accuracy. To support this research, the study introduces VAIPE, a new real-world multi-pill image dataset containing about 10,000 images of 96 different pill types. Experimental results demonstrate PGPNet's effectiveness, showing a 9.4% improvement in COCO mAP over Faster R-CNN and a 12% improvement over the YOLOv5 baseline, highlighting its potential for advancing pill identification technology.

Besides, the other paper developed a system leverages the imprinted characters on pills, which are crucial for identification, in conjunction with the pill's shape, color, and form. The model architecture integrates convolutional networks with a character-level language model. Specifically, the YOLOv5 object detection model identifies the location and types of imprinted characters on a pill image, while the ResNet-32 model captures other pill features[9]. Additionally, a recurrent neural network-based language model corrects the characters detected by YOLO using contextual information from the pill's appearance. The experimental results demonstrate that the predicted top-1 candidates achieve accuracy levels of 85.6% for the South Korean database and 74.5% for the United States database when identifying types of pills that were not included in the training set across these two different databases.

2.0 METHODOGY

The methodology of this machine learning project is designed to systematically address the challenge of pill classification using image data. The process divided into five phase which are data acquisition, model formulation, optimized method, model efficacy validation and documentation. In phase one, these process starts with the collection of

a diverse dataset from unconstrained sources, specifically Kaggle.com, known for its extensive and varied collection of images. Hyperparameters are experimentally established to each module. The dataset is divided into 20 types of pills as shown in Figure 1. The system and baseline were utilized using PyTorch and trained on a single machine equipped with OS, Win 11; CPU, 12th Gen Intel(R) Core (TM) i7-12650H@2.30 GHz, 32 GB of RAM. The experimental platform was built based on the Python programming language and the PyTorch framework. All three models were trained on this configuration. The specific parameters are shown in Table 1.



Figure 1. Images of type of pills

Table 1. Parameter configuration for all models

Parameter	Value
Batch size	8
Learning rate	0.001
Momentum	0.937
Decay	0.0005
Epochs	50

The initial step involves data pre-processing, where the raw images are oriented and resized to prepare them for annotation. Following this, the data is annotated using Labellmg software, a tool that allows for precise labeling of each pill in the images with its corresponding class as shown in Figure 2. The annotated dataset, consisting of 994 images, is then divided into 718 (70%) for training, 138 (15%) for validation, and 138 (15%) for testing. The training set was used to train the model and the validation set was used to check the state of the model during the training process to assess whether the model was over-fitting. To evaluate the system's performance, a test dataset were constructed composed of pills that were entirely different from those used in the training dataset. This approach was based on the assumption that new pills would be continually added to the database.

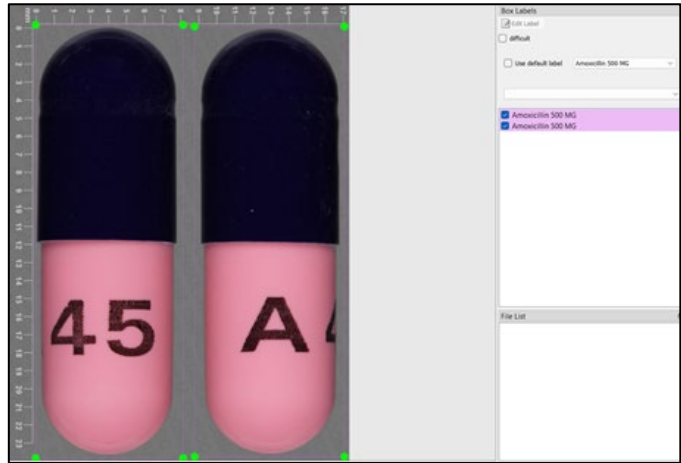


Figure 2. Image Annotation using LabelImg tool.

The next phase involves model learning, where three object detection models, YOLOv3, YOLOv5 and YOLOv8, are selected for training. These models are adopted based on Ultralytics framework. The models are employed for training using Python, with the entire process executed and documented in Jupyter Notebooks, during which the parameters are adjusted to minimize error. Throughout the training process, the models are validated on the validation dataset to monitor performance and prevent overfitting, leading to fine-tuning of hyperparameters such as learning rate and batch size.

After training and validation, the best model is selected based on its validation performance in phase three. The trained weights of this model, which represent the learned parameters, are saved and used for making predictions on the test dataset. The model is tasked with classifying the pills into 20 different categories, demonstrating its ability to differentiate between various types of medication. Phase four is the evaluation of model performance through data visualization techniques, including the creation of a confusion matrix and the calculation of precision, recall, and F1-score. These metrics provide a comprehensive understanding of the model's accuracy and efficiency in classification tasks. The entire process, from data collection to model evaluation, is thoroughly documented to ensure reproducibility and to provide insights into the methodologies and results.

In conclusion, the project demonstrates the effectiveness of using advanced object detection models for pill classification. The detailed documentation and visualization of results help in making a conclusion and suggesting potential improvements for future work. The project then ends, having achieved its goal of developing a robust model for pill classification. Figure 3 shows the flowchart of the process.

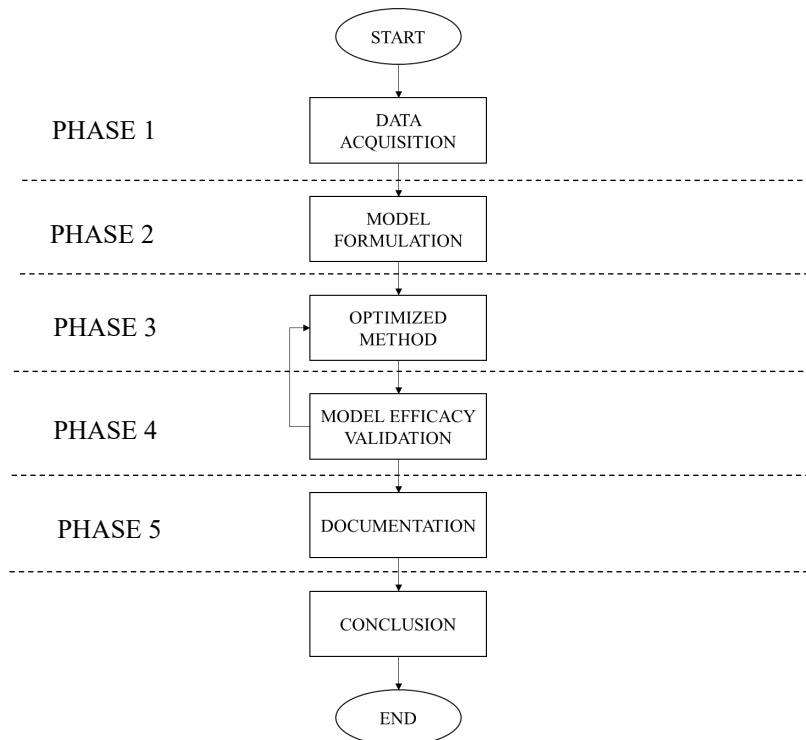


Figure 3. A flowchart of system.

3.0 RESULT AND DISCUSSION

The trained models, YOLOv3, YOLOv5 and YOLOv8, were evaluated based on their ability to classify pill images accurately. The evaluation metrics used included precision, mean Average Precision, recall, and F1-score, which provided a comprehensive understanding of the models' performance. YOLO, known for its real-time detection capabilities, and demonstrated robust performance, though with distinct strengths. Throughout the training and validation phases, certain challenges were encountered, particularly in distinguishing between pill types with similar visual features. These observations offer valuable insights into areas where the models excelled and where further improvements are needed.

3.1 COMPARISON OF ALGORITHM DETECTION RESULTS

After training, the different algorithms were utilized for pill identification on the test set, and the results are presented in Figure 4 and Figure 5. Notably, these three models can predict multiple bounding boxes and their categories simultaneously. Among them, YOLOv8 has a faster detection speed compared to the other network model structures. According to Table 2, YOLOv8 processes 8.25 images per second, whereas YOLOv3 and YOLOv5 process 0.33 images per second and 3.05 images per second, respectively. This demonstrates that YOLOv8's detection speed is significantly higher than that of YOLOv3 and YOLOv5. Based on the analysis of these experimental results, all three models are suitable for pill recognition, especially in scenarios where real-time performance is a priority and a slightly lower mAP is acceptable.

Table 2. Total time taken to process 1 image for each deep learning models

Model	Frame per second (FPS)
YOLOv3	0.33
YOLOv5	3.05
YOLOv8	8.25

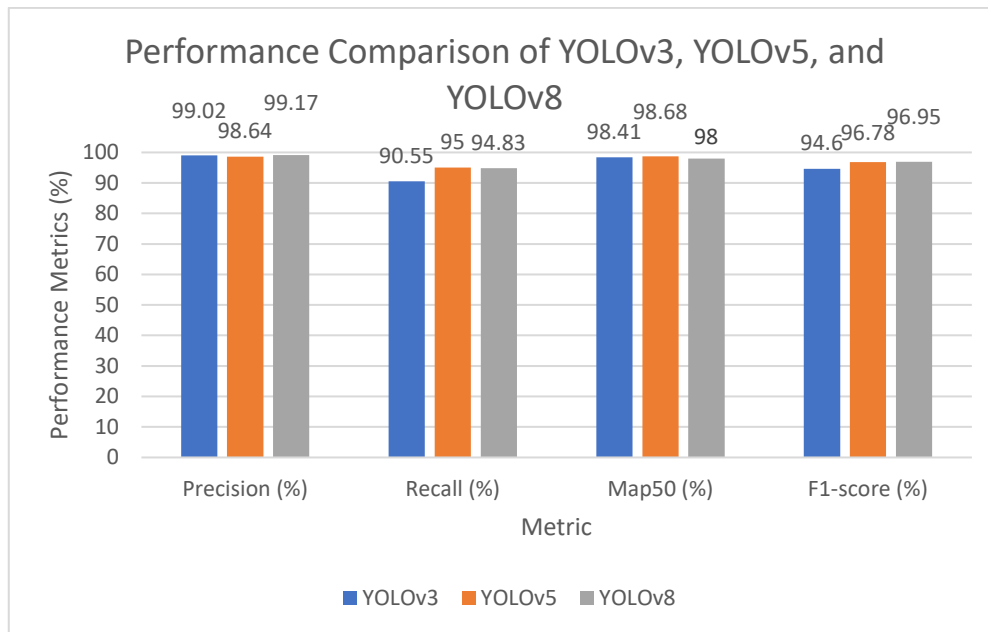


Figure 4. Histogram of evaluation of deep learning models

The performance comparison of the YOLO models (YOLOv3, YOLOv5, and YOLOv8) based on the metrics of Precision, Recall, mAP50, and F1-score reveals significant insights into their respective strengths and areas of improvement as shown in Figure 4. Precision measures the accuracy of the positive predictions made by the model. YOLOv8 shows the highest precision (99.17%), indicating that it has the lowest false-positive rate among the three models. YOLOv3 also performs well with a precision of 99.02%, while YOLOv5 is slightly behind at 98.64%. This suggests that YOLOv8 and YOLOv3 are more reliable in predicting true positives accurately. However, recall indicates the model's ability to identify all relevant instances. YOLOv5 best in this metric with a recall of 95%, followed by YOLOv8 at 94.83%. YOLOv3 lags behind with a recall of 90.55%. The higher recall values for YOLOv5 and YOLOv8 suggest that they are better at capturing all relevant instances, though YOLOv3 might miss some true positives.

Mean Average Precision at 50% (mAP50) is a comprehensive metric for assessing the accuracy of object detection models. YOLOv5 leads with the highest mAP50 of 98.68% among the models, followed by YOLOv3 at 98.41%, and YOLOv8 at 98%. These elevated mAP50 values across all three models highlight their robust performance in object detection, with YOLOv5 having a slight advantage. The F1-score, representing the harmonic mean of precision and recall, balances these two metrics. YOLOv8 achieves the highest F1-score at 96.95%, with YOLOv5 close behind at 96.78%. YOLOv3, with an F1-score of 94.6%, reflects its lower recall. The high F1-scores of YOLOv8 and YOLOv5 indicate a better balance between precision and recall compared to YOLOv3.

From Figure 5, the comparison of YOLOv3, YOLOv5, and YOLOv8 based on precision, recall, and mAP50 highlights clear differences in their performance. The precision plots for YOLOv5 and YOLOv8 show significant upward trends, stabilizing at high values with fewer fluctuations than YOLOv3. YOLOv8, in particular, exhibits the most stability and the highest precision values, indicating its superior accuracy in identifying true positives while minimizing false positives. Regarding recall, YOLOv8 again demonstrates the best performance, with its recall plot showing a steady and significant increase, stabilizing at a high value with minimal fluctuations. YOLOv5 also performs well, showing fewer fluctuations and a more consistent upward trend than YOLOv3, indicating a better ability to identify a larger proportion of actual positive cases. The mAP50 metric further underscores the superiority of YOLOv8. Its mAP50 plot shows a significant and consistent increase, stabilizing at a high value with minimal fluctuations, indicating improved overall performance in detecting objects with significant overlap with the ground truth boxes. YOLOv5 also shows better performance and stability compared to YOLOv3, with higher final mAP50 values and fewer fluctuations.

Overall, YOLOv8 outperforms in precision and F1-score, making it the top choice for accurately identifying true positives and maintaining a good balance between precision and recall. YOLOv5 stands out in recall and mAP50, showcasing its strength in detecting all relevant objects and overall accuracy in detection tasks. While still competitive, YOLOv3 generally ranks slightly below YOLOv5 and YOLOv8 on most metrics. The optimal model selection depends on specific application needs. If minimizing false positives is crucial, YOLOv8 is the best choice. If capturing all relevant instances is more important, YOLOv5 would be preferable. YOLOv3, despite its lower recall, still offers robust performance and could be considered for scenarios where computational resources or model size are constrained. Therefore, YOLO v8 has the potential to be applied to assist pharmacists to identify pills in a hospital or clinic dispensary environment as it achieved higher precision and faster in time taken on detecting object than the other two models.

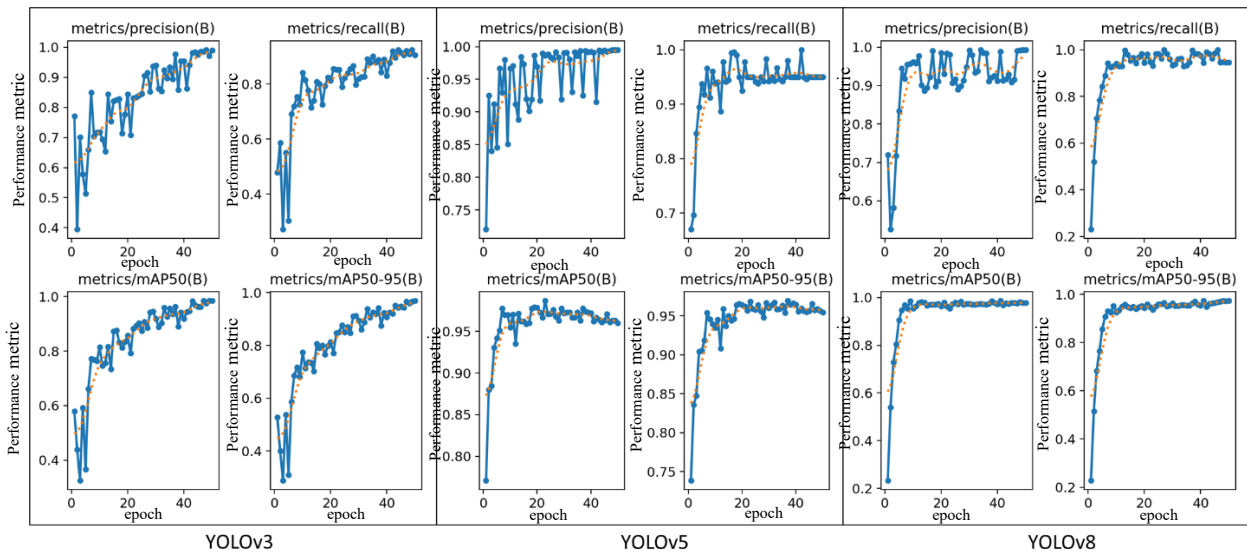


Figure 5. Model performance measures graph

4.0 CONCLUSION

Pill images were collected and utilized LabelImg to create a standard YOLO format image database. The three prominent object detection methods, YOLOv3, YOLOv5, and YOLOv8, were trained using this pill dataset. The loss function of YOLOv8 converges faster, indicating shorter training times compared to the other two models. This makes YOLOv8 more adept at handling the frequent retraining required due to frequent changes in pharmacy pill inventories. When comparing evaluation metrics, each model shows distinct strengths and weaknesses. YOLOv5 boasts a high mAP of 98.68%, but its detection speed (FPS: 3.05) lags behind YOLOv8. In high-traffic hospital pharmacies, pill identification needs both a high mAP and rapid detection speed. YOLOv8 strikes an optimal balance, offering quick and accurate drug identification, which helps reduce dispensing errors and enhance patient safety. Regarding model size, YOLOv8 is well-suited for low-performance platforms due to its fast detection speeds, making it highly applicable and valuable in practical settings. However, the study has limitations, such as a constrained experimental dataset comprised solely of split pill images sourced from a pharmaceutical online database. A larger dataset would yield more robust and real-time detection results. Additionally, training on a CPU takes significantly longer than on a GPU. In future work, a larger dataset was built and continuously test new algorithms to further optimize the model, improving both mAP and detection speed. The plan is to switch to GPU-based training to leverage its superior parallel processing and large-scale computation capabilities.

5.0 CONFLICT OF INTEREST

Following best practices, the authors have transparently disclosed all financial and non-financial interests that might influence the manuscript's content. Importantly, they affirm that no conflicts of interest exist.

6.0 AUTHOR CONTRIBUTION

Mohd Rais Hakim Bin Ramlee: Responsible for methodology, validation, formal analysis, data curation, investigation, resources, software, visualization, and writing the original draft.

Ismail Mohd Khairuddin: Oversaw supervision, funding acquisition, review and editing of the manuscript, and project administration.

Muhammad Amirul Abdullah: Oversaw supervision.

Zubaidah Zamri: Contributed to writing.

Muhammad Nur Aiman Shapiee: Handled software and visualization.

7.0 ACKNOWLEDGEMENT

The authors express their sincere gratitude to Universiti Malaysia Pahang Al-Sultan Abdullah for providing funding for this study through RDU223231.

8.0 REFERENCES

- [1] S. S. Kar, H. S. Pradhan, and G. P. Mohanta, "Concept of essential medicines and rational use in public health," *Indian J. Community Med.*, vol. 35, no. 1, pp. 10–13, 2010, doi: 10.4103/0970-0218.62546.
- [2] S. Sultana, "History of Contraception," *J. Islam. Int. Med. Coll.*, vol. 14, no. 4, pp. 172–173, 2019, doi: 10.32440/ar.2018.135.01.rev09.
- [3] M. Janczura, S. Sip, and J. Cielecka-Piontek, "The Development of Innovative Dosage Forms of the Fixed-Dose Combination of Active Pharmaceutical Ingredients," *Pharmaceutics*, vol. 14, no. 4, 2022, doi: 10.3390/pharmaceutics14040834.
- [4] E. Nabrawi and A. T. Alanazi, "Imaging in Healthcare: A Glance at the Present and a Glimpse Into the Future," *Cureus*, vol. 15, no. 3, 2023, doi: 10.7759/cureus.36111.
- [5] C. O. Airhihenbuwa, T. S. Tseng, V. D. Sutton, and L. S. Price, "Global Perspectives on Improving Chronic Disease Prevention and Management in Diverse Settings," *Prev. Chronic Dis.*, vol. 18, no. December 2020, pp. 1–7, 2021, doi: 10.5888/PCD18.210055.
- [6] Y. Zheng *et al.*, "Designing Human-Centered AI to Prevent Medication Dispensing Errors: Focus Group Study With Pharmacists," *JMIR Form. Res.*, vol. 7, no. 1, pp. 1–15, 2023, doi: 10.2196/51921.
- [7] C. Chakraborty, M. Bhattacharya, S. Pal, and S. S. Lee, "From machine learning to deep learning: Advances of the recent data-driven paradigm shift in medicine and healthcare," *Curr. Res. Biotechnol.*, vol. 7, no. November 2023, p. 100164, 2024, doi: 10.1016/j.crbiot.2023.100164.
- [8] A. R. Ashraf, A. Somogyi-Végh, S. Merczel, N. Gyimesi, and A. Fittler, "Leveraging code-free deep learning for pill recognition in clinical settings: A multicenter, real-world study of performance across multiple platforms," *Artif. Intell. Med.*, vol. 150, no. June 2023, 2024, doi: 10.1016/j.artmed.2024.102844.
- [9] J. Heo, Y. Kang, S. K. Lee, D. H. Jeong, and K. M. Kim, "An Accurate Deep Learning-Based System for Automatic Pill Identification: Model Development and Validation," *J. Med. Internet Res.*, vol. 25, 2023, doi: 10.2196/41043.
- [10] E. Y. Fung, B. Leung, D. Hamilton, and J. Hope, "Do automated dispensing machines improve patient safety?,"

- Can. J. Hosp. Pharm.*, vol. 62, no. 6, pp. 516–519, 2009, doi: 10.4212/cjhp.v62i6.852.
- [11] S. S. Gomasta, A. Dhali, T. Tahlil, M. M. Anwar, and A. B. M. S. Ali, “PharmaChain: Blockchain-based drug supply chain provenance verification system,” *Heliyon*, vol. 9, no. 7, p. e17957, 2023, doi: 10.1016/j.heliyon.2023.e17957.
- [12] S. Walker, “Foreword by Professor Stuart Walker,” *Int. Coop. Converg. Harmon. Pharm. Regul.*, pp. xvii–xviii, 2014, doi: 10.1016/b978-0-12-800053-3.06001-8.
- [13] J. Kufel *et al.*, “What Is Machine Learning, Artificial Neural Networks and Deep Learning?—Examples of Practical Applications in Medicine,” *Diagnostics*, vol. 13, no. 15, 2023, doi: 10.3390/diagnostics13152582.
- [14] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.
- [15] Y. Zhu, R. Urtasun, R. Salakhutdinov, and S. Fidler, “SegDeepM: Exploiting segmentation and context in deep neural networks for object detection,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 4703–4711, 2015, doi: 10.1109/CVPR.2015.7299102.
- [16] L. Alzubaidi *et al.*, *Review of deep learning: concepts, CNN architectures, challenges, applications, future directions*, vol. 8, no. 1. Springer International Publishing, 2021. doi: 10.1186/s40537-021-00444-8.
- [17] S. B. Junaid *et al.*, “Recent Advancements in Emerging Technologies for Healthcare Management Systems: A Survey,” *Healthc.*, vol. 10, no. 10, 2022, doi: 10.3390/healthcare10101940.
- [18] A. Salam and N. Abhinesh, “Revolutionizing dermatology: The role of artificial intelligence in clinical practice,” *IP Indian J. Clin. Exp. Dermatology*, vol. 10, no. 2, pp. 107–112, 2024, doi: 10.18231/j.ijced.2024.021.
- [19] P. T. Wu, T. Y. Sun, J. C. Lin, and C. L. Chin, “Round Pill Shape Recognition System Based on AY Deep Learning Model,” in *Proceedings - 2022 IEEE International Conference on Consumer Electronics - Taiwan, ICCE-Taiwan 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 553–554. doi: 10.1109/ICCE-Taiwan55306.2022.9869124.
- [20] L. Tan, T. Huangfu, L. Wu, and W. Chen, “Comparison of RetinaNet, SSD, and YOLO v3 for real-time pill identification,” *BMC Med. Inform. Decis. Mak.*, vol. 21, no. 1, Dec. 2021, doi: 10.1186/s12911-021-01691-8.
- [21] A. D. Nguyen, H. H. Pham, H. T. Trung, Q. V. H. Nguyen, T. N. Truong, and P. Le Nguyen, “High accurate and explainable multi-pill detection framework with graph neural network-assisted multimodal data fusion,” *PLoS One*, vol. 18, no. 9 September, Sep. 2023, doi: 10.1371/journal.pone.0291865.