Pharmaceutical Image Categorization

1st Shireesh Kumar Poral Ashok Kumar
Department of Computer Science
University of Houston
Texas, USA
sporalas@cougarnet.uh.edu

2nd Maria Ann Toms
Department of Computer Science
University of Houston
Texas, USA
mtoms@cougarnet.uh.edu

3rd Timothy A Burt
Department of Physics
University of Houston
Texas, USA
taburt@central.uh.edu

4th Ioannis A Kakadiaris
Department of Computer Science
University of Houston
Texas, USA
ikakadia@central.uh.edu

***Abstract*—In this study, we address the challenge of detecting counterfeit pharmaceutical products sold online through advanced image classification techniques. Initially, we employed a dataset of AI-generated images categorized by risk levels to train models such as the Vision Transformer (ViT) and ResNet-50. However, due to inadequate initial results, the approach was revised to utilize a more robust, manually curated dataset comprising images scraped from Google, covering a range of pharmaceutical products. We explored several architectures, ultimately selecting the SWIN Transformer over the ViT due to its hierarchical structure, efficiency in handling shifted window mechanisms, and superior scalability. The final model incorporated additional fully connected layers and employed a dual-classification strategy to distinguish whether an image represents a medical product and to classify its specific category. This dual approach significantly enhanced the model's precision in identifying potential counterfeit products. Our findings demonstrate the potential of using sophisticated image classification methods to assist in regulatory and safety monitoring.**

***Index Terms*—Counterfeit Pharmaceuticals, Image Classification, SWIN Transformer, Vision Transformer, Deep Learning, Online Pharmacy Safety, AI in Pharmaceutical Regulation**

I. Introduction

The global increase in online sales of pharmaceutical products has brought with it a parallel rise in the distribution of counterfeit medications, posing significant challenges to public health and safety. Counterfeit pharmaceuticals often contain harmful ingredients, incorrect dosages, or may entirely lack the purported active substances, leading to ineffective treatment or severe health complications. As these counterfeit products become more sophisticated, the ability to identify and eliminate them from circulation becomes both increasingly complex and crucial.

The urgency of addressing this issue is underscored by the potential life-threatening risks associated with the consumption of counterfeit drugs. These risks include the possibility of poisoning, allergic reactions, and the worsening of medical conditions due to ineffective treatment. Furthermore, the presence of counterfeit medications undermines public trust in healthcare systems and pharmaceutical supply chains, exacerbating the challenges faced by public health authorities worldwide.

This research focuses on harnessing advanced image classification technologies to tackle the problem of counterfeit pharmaceuticals effectively. By leveraging deep learning models like the ResNet-50 and the Vision Transformer (ViT), our initial studies aimed to develop a robust system capable of identifying counterfeit pharmaceuticals through visual analysis. However, the initial experiments using AI-generated images and traditional deep learning models revealed limitations in handling the complexity and variability of real-world data. This insight led to a pivot toward utilizing the SWIN Transformer, an architecture renowned for its superior performance in hierarchical processing of visual information.

The main objectives of this system are twofold: firstly, to determine the authenticity of medical products based on their visual characteristics and secondly, to enhance the precision and reliability of these classifications to support the efforts of regulatory bodies and pharmaceutical companies. By developing a model that can accurately classify and identify counterfeit pharmaceuticals, we aim to contribute significantly to safeguarding public health and maintaining the integrity of global pharmaceutical markets.

This paper outlines the development process of our classification system, presents the results from extensive validations, and discusses the potential impact and future directions of our research. Through this work, we demonstrate the critical role of advanced machine learning techniques in combating the global issue of counterfeit pharmaceuticals sold online.

II. Background

The rise of e-commerce has revolutionized the way pharmaceutical products are bought and sold, providing consumers with unparalleled access and convenience. However, this digital transformation has also created new opportunities for the sale and distribution of counterfeit medications [1]. Counterfeit pharmaceuticals pose significant risks to public health, as they often contain incorrect doses, harmful ingredients, or no

active ingredients at all [2]. The World Health Organization estimates that up to 10% of drugs worldwide are counterfeit, with even higher percentages in developing countries [3]. To combat this growing threat, researchers have turned to machine learning and computer vision techniques to develop automated systems for detecting counterfeit products [4]. Convolutional Neural Networks (CNNs) have been widely used for this purpose due to their ability to learn hierarchical features from images [5]. For example, a study by Dey et al. used a CNN to classify genuine and counterfeit pills, achieving an accuracy of 97.5% [6]. Similarly, Vajda et al. employed a CNN to detect counterfeit packaging of medical products, reporting an accuracy of 96.7% [7]. More recently, transformer-based architectures, such as the Vision Transformer (ViT) [8] and the SWIN Transformer [9], have emerged as powerful alternatives to CNNs for image classification tasks. These models leverage self-attention mechanisms to capture global dependencies and have shown impressive results on various benchmarks. However, their application to the domain of counterfeit pharmaceutical detection remains largely unexplored. In this study, we aim to address this gap by investigating the use of advanced transformer-based models for identifying counterfeit pharmaceutical products sold online. By leveraging the hierarchical structure and shifted window mechanism of the SWIN Transformer, we develop a robust system capable of accurately classifying medical products into specific categories. Our work builds upon previous research in the field while introducing novel techniques to enhance the precision and reliability of counterfeit detection systems.

III. Dataset Description

This section details the datasets used for our experiments, which evolved from an initial unsuccessful attempt to a redefined approach with better-defined criteria.

A. Initial Experiment

Initially, we attempted to implement an AI model using generated images of pharmaceutical products commonly sold online and susceptible to counterfeiting. The product categories included:

- Asthma Inhalers
- COVID Test Kits
- Face Masks
- Medicine Bottles
- Ointments
- Syringes
- Tablet Sheets

Each category had sub-folders labeled as high, medium, and low risk, with each containing 10 images. High-risk product images contained quantities greater than 50, medium risk between 10 and 50, and low risk fewer than 10. The risk levels were initially defined by quantity and used as prompts in Google Gemini for image generation. However, this definition of risk was found to be inappropriate for our goals, leading to a reevaluation of the project and dataset.

B. Revised Dataset

After revising our approach, we scraped images from Google for the following new categories, each chosen because they contain identifiable medical ingredients, aiding in the detection of counterfeit products sold online:

- Blister Packs
- Bottle Medicines
- Box Images Only
- Noise
- Ointments
- Packets of Sachets
- Powdered Medicines
- Prefilled Syringes

Each category contains 300 images. We allocated 20% of the total images for testing. From the remaining, 25% were used for validation and the rest for training. All images were manually scraped from Google, and credit goes to the original image owners.

Specific Categorization Rules:

- In the ”Box Images Only” category, only images of boxes are included. For other categories, while the primary class-specific item is the focus, boxes may be present but should not dominate the image or include any overlapping items from other classes
- An image is assigned to a specific category only if the item from that category is clearly visible and not inside a box. If it is inside a box and not visible, it is categorized under ”Box Images Only.” If the item is visible even within a box, it is placed in the appropriate separate category
- The ”Raw Powder” category was specifically included to assist investigators in identifying illegal sales of raw powders online

Note: This dataset is solely used for academic purposes and all images are credited to their respective owners.

IV. Architecture Overview

In our research, we explored various deep learning architectures to best address the challenges posed by image classification of pharmaceutical products. Our investigation initially included the Vision Transformer (ViT) and ResNet-50 models. However, these models were later set aside in favor of more advanced architectures due to specific project needs.

A. Initial Model Exploration

Initially, the ViT model, which applies the mechanisms of transformers, primarily used in NLP, to image processing, was paired with the traditional ResNet-50, a convolutional neural network known for its effectiveness in deep residual learning. While both models performed adequately, the ViT model struggled with computational efficiency and required substantial amounts of data to surpass the more conventional CNNs like ResNet-50. This led us to explore alternative architectures that could offer better performance with our dataset’s constraints.

B. Selection of SWIN Transformer

We ultimately selected the SWIN Transformer as our primary architecture for several reasons:

- **Hierarchical Structure:** Unlike ViT, which processes images in fixed-size patches uniformly across various scales, SWIN Transformer constructs a hierarchical representation that allows it to adapt more effectively to various image sizes and resolutions. This is particularly beneficial for our dataset, which includes diverse image categories
- **Shifted Windows:** The SWIN Transformer introduces shifted windows, which limit self-attention computation to non-overlapping local windows while also allowing for cross-window connection. This design significantly reduces the computational burden compared to the global self-attention mechanism in ViT
- **Scalability:** SWIN Transformers are inherently more scalable across different model sizes and are more adaptable to different image resolutions, which enhances their usability in practical applications beyond academic research
- **Better Performance:** In practice, SWIN Transformers have demonstrated superior performance on benchmarks involving image classification, object detection, and semantic segmentation compared to their predecessors

C. Configuration in Main Experiment

In the main experiment, both the Vision Transformer (ViT) and SWIN Transformer were enhanced with additional fully connected (FC) layers to improve their learning capacity for specific tasks. The FC layers added were 2048, 1024, 512 units respectively, followed by the output layer corresponding to the number of classes. This configuration was intended to refine the feature extraction capabilities before the final classification.

Furthermore, we employed a dual-classifier setup for the later part of our experiments:

- The first classifier determines whether the image represents a medical product
- The second classifier identifies the specific category to which the medical product belongs

This approach allowed us to effectively manage the multilayer classification challenges posed by our diverse dataset, enhancing both accuracy and specificity in identifying counterfeit pharmaceutical products.

D. Figure Inclusion

Note: The image is used for educational purposes, and all rights belong to the original creator.

V. Experimentation

A. Initial Experimentation

In the initial phase of our study, we employed two distinct models to evaluate the performance on a dataset of AI-generated pharmaceutical images: the ResNet-50 and the Vision Transformer (ViT). We chose ResNet-50 to gauge the effectiveness of convolutional neural networks (CNNs) on synthetic data. For a broader perspective involving transformer-based models, we tested the ViT with different configurations, specifically ViT with 16-patch and 32-patch sizes, to compare their efficiencies.

In our implementation of the SWIN Transformer, we opted to utilize a simplified version of the architecture, focusing on a single stage (or transformer block) of the model. This decision was driven by [reasons such as computational constraints, dataset characteristics, etc.]. While this does not exploit the full hierarchical capability of the SWIN architecture, it provides sufficient performance for our specific application needs, balancing complexity and computational efficiency.

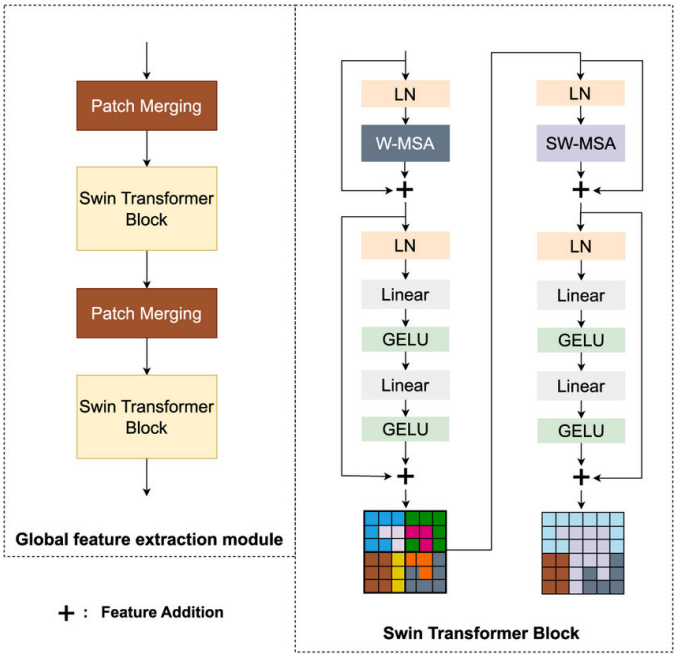

Fig. 1. SWIN Transformer architecture.

B. Refined Experimentation

Based on our findings from the initial tests, we continued with the ViT 16-patch model for the refined experiment, as it outperformed the 32-patch configuration in terms of accuracy and computational efficiency. Additionally, we incorporated the SWIN Transformer Base Model, chosen for its superior hierarchical structure and shifted window mechanism, which is more suited for our complex image datasets.

In this main phase of experimentation, we focused on two primary classification tasks:

- Binary classification to determine if an image is related to medical products
- Class-specific classification to identify the exact category of the medical product depicted in the image

Both models were trained for approximately 50 epochs. We initially utilized the pretrained ImageNet-1K weights, which were then frozen to maintain the integrity of learned features. Subsequent training phases involved fine-tuning on

the additional fully connected (FC) layers that were introduced to cater to the specific needs of our dataset and research focus. Additionally, we fine-tuned the learning rate and regularization parameters (lambda) to optimize performance and prevent overfitting during experimentation.

VI. Applications and Uses

The developed image classification system has several practical applications across different sectors, enhancing both safety and compliance. Here are some key applications:

A. Proactive Brand Protection

Pharmaceutical companies can use this technology to monitor online platforms for counterfeit products actively. By quickly identifying fake products, companies can take swift action to protect their brand reputation and ensure consumer safety, ultimately safeguarding their market share and public trust.

B. Law Enforcement and Regulatory Support

This system aids law enforcement agencies and regulatory bodies by streamlining the investigation process. It helps in quickly identifying counterfeit or unauthorized medical products, thereby facilitating faster and more accurate legal actions. This contributes significantly to improving public health outcomes by removing potentially harmful counterfeit medications from the market.

C. Enhanced Surveillance and Monitoring

The technology enables continuous surveillance of online pharmaceutical sales, providing real-time alerts on suspicious activities. This helps in pre-emptive action against the distribution of illegal medications and supports public health officials in crisis management and disease prevention.

D. Integration with OCR for Detailed Analysis

Once the system identifies the type of medical product and its packaging, it can integrate with Optical Character Recognition (OCR) technology to extract and analyze textual information from the product's image. This can be crucial for verifying product details such as batch numbers, expiration dates, and manufacturer information, further aiding in authenticity verification.

E. Educational and Training Tool

The system can also serve as an educational and training tool for pharmacists, customs officers, and healthcare professionals. By familiarizing them with the characteristics of counterfeit drugs, they can better identify and report such items, enhancing overall healthcare safety.

VII. Results

A. Initial Experiments with ViT and ResNet

Our initial experiments involved training the Vision Transformer (ViT) with different patch sizes and epochs, as well as using the traditional ResNet-50 model. The results demonstrated varying degrees of success:

- The ViT model with 16 patches reached an accuracy of 83% on the test set after 20 epochs and improved to 89% after 40 epochs
- The 32-patch ViT model showed no improvement with increasing epochs, maintaining an accuracy of 69% both after 20 and 40 epochs
- ResNet-50, after 50 epochs, significantly underperformed in comparison, achieving only 45% accuracy on the test set

Loss plots for training and validation, along with confusion matrices for these models, provide further insights into their performance and are included below.

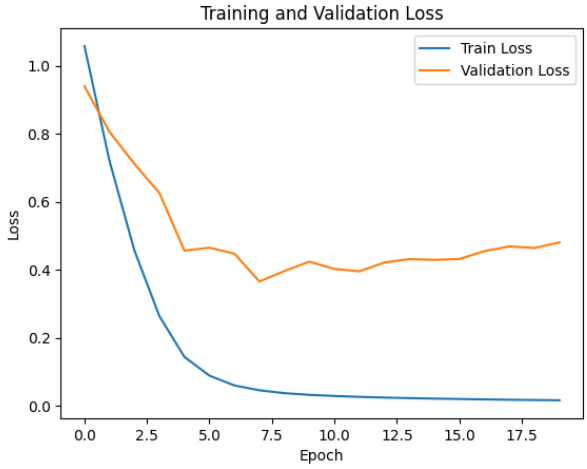

Fig. 2. Training and Validation Loss Plot for ViT 16 patch (40 epochs training)

B. Advanced Experiments with SWIN Transformer and ViT 16 Patch

In our advanced experiments, we focused on two configurations: the SWIN Transformer base model and a refined ViT 16 patch model, assessing their performance in both binary (medical vs. non-medical) and multi-class (7 classes) settings:

- The SWIN Transformer base model achieved an accuracy of 79% for the binary classification and 94% for the 7-class classification
- The ViT 16 patch showed an improvement in the binary classification to 82% but a decrease in the 7-class classifier accuracy to 23%
- The classification categories are as follows: Class 0 represents Blister Packs, Class 1 represents Bottles of Medicine, Class 2 represents Boxes, Class 3 represents Noise, Class 4 represents Ointments, Class 5 represents

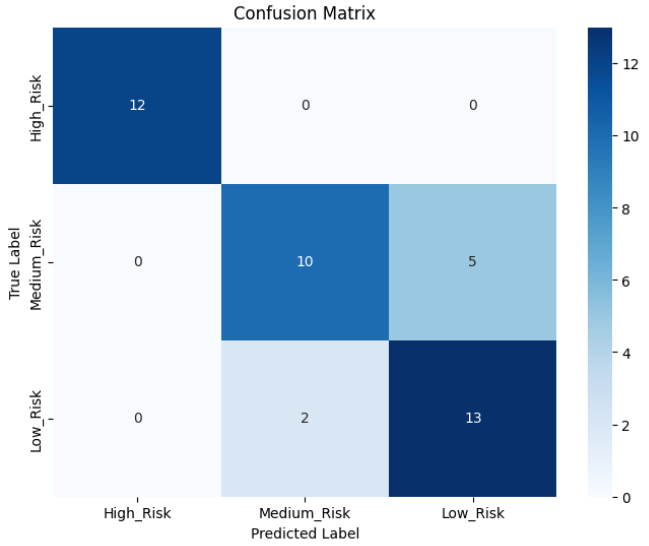

Fig. 3. Confusion Matrices for Testing ViT 16 patch (40 epochs training)

Packets, Class 6 represents Powder Medications, and Class

Similar to our initial experiments, we provide corresponding loss plots and confusion matrices for these tests to detail the model performances.

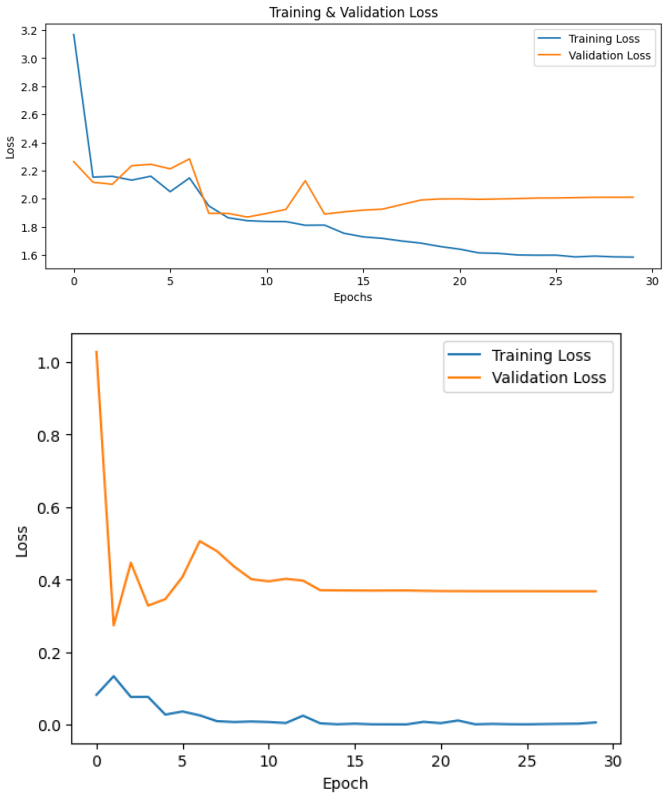

Fig. 4. Training and Validation Loss Plots for SWIN Transformer Base Model: (A) Binary Classifier, (B) 7-class classifier

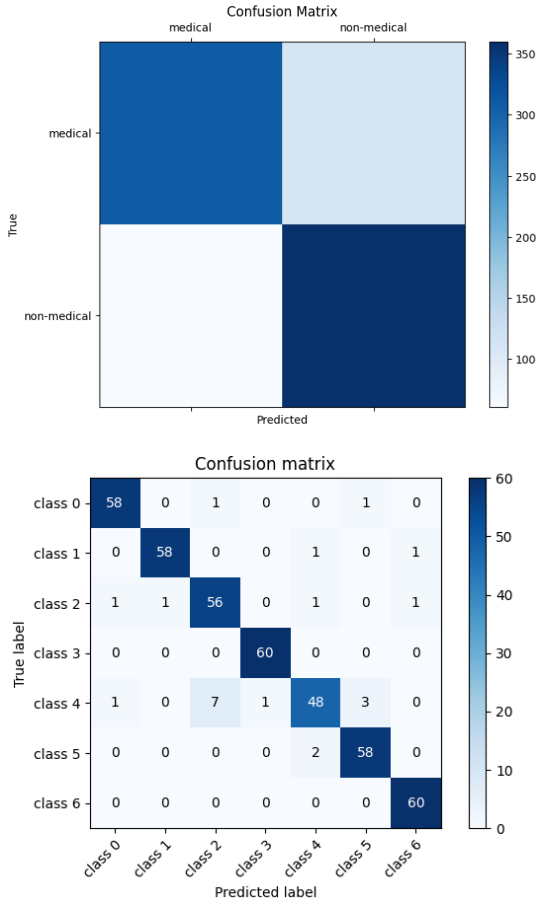

Fig. 5. Confusion Matrices for Testing SWIN Transformer Base Model: (A) Binary Classifier, (B) 7-class classifier.

C. Discussion

The results underscore the challenges and complexities involved in training deep learning models for pharmaceutical image classification. The SWIN Transformer's performance suggests it may be better suited for handling the nuances of medical product imagery compared to the ViT models. Further analysis and optimization could enhance these models' efficacy, particularly in the application of real-world scenarios.

VIII. Future Work

The progress achieved in this project opens several avenues for future development, aiming to enhance the system's efficacy and applicability in real-world scenarios. Below are key areas of focus for future work:

A. Optimization and Accuracy Enhancement

Future efforts will concentrate on optimizing the current model to enhance its accuracy and efficiency. This involves refining the algorithms and employing advanced machine learning techniques to improve the model's ability to discern and classify medical products with greater precision. Techniques such as hyperparameter tuning, advanced regularization methods, and ensemble learning could be explored to achieve these goals.

B. Integration of OCR for Text Extraction

To expand the model's capabilities, we plan to integrate Optical Character Recognition (OCR) technology. This will allow the system to extract and analyze textual information from images of medical products, such as drug names, dosages, and expiration dates. Incorporating OCR will provide additional data points that can significantly enhance the accuracy of the classification models and assist in more detailed verification processes.

C. Real-time Processing Capabilities

Developing the model to function effectively in real-time applications is crucial. This involves optimizing the model to reduce latency and increase throughput, ensuring that it can operate within the constraints of real-time systems. Techniques such as model quantization, pruning, and deploying on specialized hardware like FPGAs or TPUs might be necessary to achieve real-time performance.

D. Adaptive Learning for Novel Threats

To keep the model relevant in the face of evolving counterfeit strategies, incorporating adaptive learning mechanisms will be critical. This would allow the model to continually learn from new data, adapting to new types of counterfeits and potentially unseen medical products without requiring complete retraining.

E. Enhancing Model Robustness

Further work will also focus on enhancing the robustness of the model against various forms of adversarial attacks and common image distortions encountered in real-world scenarios. Implementing robustness tests and adjusting the model to maintain high performance under different environmental conditions and attack vectors will be essential for practical deployment.

These enhancements will aim to make the model not only more accurate but also more versatile and applicable in a range of real-time settings, thus contributing to safer pharmaceutical practices globally.

IX. Conclusion

This study has demonstrated the feasibility and effectiveness of using advanced image classification models to combat the proliferation of counterfeit pharmaceutical products online. Our dual-classification system was specifically developed to discern not only the authenticity of medical products but also to accurately categorize them into specific classes based on their visual characteristics.

The medical vs. non-medical classifier achieved a commendable accuracy of 79%, providing a robust foundation for initial screening of pharmaceutical products. However, the more focused classifier, which categorizes medical products into specific classes, exhibited superior performance with an impressive accuracy of 94%. This high level of precision is particularly significant as it directly supports the main application of our project—ensuring that medical products can be accurately and reliably authenticated and classified in real-time.

The results from this study highlight the potential for deep learning technologies to significantly enhance the security and reliability of pharmaceutical distributions in online platforms. By continuing to refine these models, particularly by focusing on improving the accuracy of the initial medical vs. non-medical classification stage, we can further strengthen the tool's utility and reliability.

Future work will focus on expanding the dataset and incorporating more varied examples to challenge the model's robustness and accuracy. Additionally, continuous optimization of the algorithms and enhancement of the system's ability to adapt to new and evolving types of counterfeit products will be crucial.

In conclusion, the developed system stands as a promising tool in the ongoing effort to protect consumers from counterfeit pharmaceuticals and support regulatory bodies in maintaining the integrity of global health systems. By providing a means to quickly and accurately screen and classify pharmaceutical products, this technology can play a pivotal role in enhancing public health safety and strengthening consumer trust in pharmaceutical markets.

References

[1] Mackey, T.K., Nayyar, G.: A review of existing and emerging digital technologies to combat the global trade in fake medicines. *Expert opinion on drug safety*, 1–8 (2018)

[2] Blackstone, E.A., Fuhr Jr, J.P., Pociask, S.: The health and economic effects of counterfeit drugs. *American health drug benefits **7(4)***, 216 (2014)

[3] World Health Organization: WHO Global Surveillance and Monitoring System for substandard and falsified medical products. Tech. rep., World Health Organization (2017)

[4] Rasheed, A., Ajmal, M., Qayyum, A., Khelifi, A., Qureshi, R., Jalal, A., Mahmood, M.T., Khan, R.: Deep learning for detection and classification of pharmaceutical packaging and counterfeit medicines. *Multimedia Tools and Applications **80(15)***, 22975–23001 (2021)

[5] LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature **521(7553)***, 436–444 (2015)

[6] Dey, S., Shi, Y., Li, Y., Zhou, H., Liu, C., et al.: Deep learning-based pill image recognition. In: *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pp. 2705–2708 (2019). IEEE

[7] Vajda, S., Lednev, K., Pollner, P., Griff, A.M., Saha, S., Ayoub, A., Al-Hashimi, H.M., Mohiuddin, A., et al.: Accurate identification of counterfeit medicines using convolutional neural networks. In: *2018 IEEE International Conference on Big Data (Big Data)*, pp. 5021–5023 (2018). IEEE

[8] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020)

[9] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. *arXiv preprint arXiv:2103.14030* (2021)