FMDB *Transactions on Sustainable Computing Systems*

Advanced Hybrid Model for Precise Automated Pharmaceutical Pill Detection and Identification Systems

Kumudha Madesh[1,*], Kannammal Ayyasamy[2], Suresh Chockkan Chetty[3], Vinitha Thangavel[4], S. J. Vimal Aravintha[5], Raja Brahmendra Chowdary Veerepalli[6], Adithi Venkatakrishnan[7]

[1,3]Department of Computer Science and Engineering, Jayalakshmi Institute of Technology, Dharmapuri, Tamil Nadu, India.
[2,4]Department of Information Technology, Jayalakshmi Institute of Technology, Dharmapuri, Tamil Nadu, India.
[5]Department of Data Science, Indiana University Bloomington, Bloomington, Indiana, United States of America.
[6]Department of Computer Science, Texas State University, San Marcos, Texas, United States of America.
[7]Department of Information Technology and Management, University of Texas at Dallas, Richardson, Texas, United States of America.
km1913511@gmail.com[1], kannammal.cse1@gmail.com[2], sureshatmc@gmail.com[3], vinitha104100@gmail.com[4], vimasubr@iu.edu[5], kgm120@txstate.edu[6], dal076222@utdallas.edu[7]

Abstract: The pharmaceutical sector is experiencing some serious problems in terms of drug mistakes and the emergence of fake medications. This study presents a powerful hybrid architecture suitable for detecting and identifying medicinal pills with high accuracy using a multi-stage deep learning framework. This research uses a dataset of 479 high-resolution data points that include different shapes, colors and pill imprints. The system combines Convolutional Neural Networks, which extract features and Transformer-based networks, which recognize sequences and patterns, resulting in improved accuracy across different lighting conditions. The software used includes Python-based deep learning libraries, specialized image augmentation frameworks and hardware-based processing units required to deliver real-time performance. The findings indicate that combining spatial and contextual information greatly reduces misidentification rates compared with single models. This mixed design offers a scalable, automated pharmacy check system and a patient-side safety app that ensure high fidelity between physical and electronic prescriptions.

Keywords: Pharmaceutical Recognition; Feature Extraction; Medication Safety; Pharmaceutical Interventions; Healthcare Technology; Computer Vision; Hybrid Learning; Computer Vision.

Received on: 19/06/2025, **Revised on:** 14/08/2025, **Accepted on:** 11/11/2025, **Published on:** 10/08/2026

Journal Homepage: https://www.fmdbpub.com/user/journals/details/FTSCS

DOI: https://doi.org/10.69888/FTSCS.2026.000721

Cite as: K. Madesh, K. Ayyasamy, S. C. Chetty, V. Thangavel, S. J. V. Aravintha, R. B. C. Veerepalli, and A. Venkatakrishnan, "Advanced Hybrid Model for Precise Automated Pharmaceutical Pill Detection and Identification Systems," *FMDB Transactions on Sustainable Computing Systems*, vol. 4, no. 3, pp. 187–195, 2026.

1. Introduction

[*]Corresponding author.

The rapid advancement of healthcare technology has created the need for automated mechanisms to verify oral solid dosage forms, as illustrated by the smart healthcare automation frameworks developed in the previous study. Medication errors are among the major causes of avoidable harm in clinical practice and are frequently associated with visual similarities among various pharmaceutical products, a risk factor in previously conducted patient safety studies [12]. Pill identification is a complex process, as numerous drugs share the same colour palette, round shapes, or minimalist debossed patterns, as discussed in pharmaceutical recognition systems developed by scientists. Conventional manual checks are susceptible to human error and control, especially in high-volume settings such as hospital pharmacies, as noted in past workflow-efficiency research. Thus, the key to current therapeutic integrity is a computerized system capable of differentiating between look-alike and sound-alike drugs, as stressed in medication assurance models proposed in previous literature [14]. This paper presents a hybrid approach that treats pill identification as both a visual classification task and a structural analysis task, thereby considering all physical features at the detection stage, motivated by multimodal systems developed by researchers.

Advanced computer vision methods have provided a basis for addressing these complexities, as confirmed by recent intelligence studies of images [9]. But single-model architectures tend to struggle with occlusions, shadows, and reflective surfaces, such as those found in blister packaging, a problem identified in earlier imaging robustness studies [5]. Our study overcomes these shortcomings by adopting a two-stream processing pipeline. The former stream considers the macro-properties of the pill, including main dimensions and prevailing colour frequencies, which are used in macro-feature extraction models [11]. The second stream examines micro-phenomena, such as the coating texture and the accuracy of imprinted alphanumeric codes, a technique investigated in fine-grained recognition systems. The hybrid model uses a synthesis of these two streams of data to form a complete digital fingerprint of each medication, as shown in fusion architectures of previous researchers [15]. Not only does this method increase detection accuracy, but it also increases the system's reliability across a wide range of hardware platforms, including industrial scanners and mobile device cameras, a benefit observed in cross-platform deployment research. In addition to clinical correctness, such systems have significant implications for the security of global supply chains, as seen in anti-counterfeit pharmaceutical models [13]. The emergence of fake drugs requires a verification process that is hard to subvert, and this can be achieved through safe validation systems developed by researchers.

The proposed hybrid model can detect inconsistencies that the eye cannot by referring to specialized topographic maps of pill surfaces, as demonstrated in previous surface-mapping methods [8]. This study combines classical image processing with recent neural networks and seeks to provide a smooth interface between physical objects and online databases. The overall goal is to design a framework that is highly sensitive and low in false positives, thereby building trust in automated pharmaceutical interventions. The sections below describe the development of these technologies and the methodological decisions that characterize this hybrid strategy. This model can operate under various conditions thanks to on-device processing and cloud-native synchronization. When specialized pharmacy personnel are unavailable in a particular area, the automated identification tool acts as an important safety net. This study, based on 479 data points, offers a proven concept for a computationally efficient, highly accurate system. The following discussion focuses on the process of raw image capture through final classification, and the significance of data residency and localized processing in preserving patient privacy and pharmaceutical accuracy. Through testing and architectural innovation, the study will add an essential element to the overall picture of intelligent healthcare infrastructure and autonomous medication management systems.

2. Review of Literature

Recent studies in pharmaceutical imaging have focused more on structural analysis than on basic colour matching, as noted in contemporary imaging studies pioneered by earlier researchers. Initial experiments in pill identification relied on edge detection and colour histograms, which were often insufficient in the face of changes in ambient light or noise, as was the case with some of the earliest recognition systems [11]. These initial approaches highlighted the need for more robust feature descriptors that remain invariant under geometric transformations, as shown in earlier research on feature engineering [2]. Academics have investigated the application of local binary patterns to reproduce the recognizable textures of medicinal pill coatings, finding that surface roughness is a good predictor of whether a medication is generic or branded, an understanding established in the literature on texture in pharmaceuticals [14]. This literature highlights the importance of high-resolution data collection to reveal fine details of tablet topography, underscoring the need for precision in imaging models [7]. The introduction of deep learning was a major turning point in the literature, and numerous studies have reported significant improvements in classification benchmarks, as seen in neural diagnostic systems proposed by previous researchers [1]. Residual networks have been successfully applied to the problem of deep layers needed in the feature hierarchy, a structure present in research on deep feature extraction [10].

These works tend to draw attention to the challenge of learning with small samples, since pharmaceutical data is often unbalanced, with millions of drugs on the market, an issue discussed in the dataset optimization literature [13]. In response, rotation, scaling, and colour jittering (augmentation strategies) have become the norm and are used in data enrichment models constructed in the past [4]. According to the literature, deep models are also very effective at global feature recognition but

occasionally fail to detect localized imprints, as observed in comparative learning tests [9]. This has prompted the need to explore attention mechanisms that can be directed to the debossed part of a pill, where important identification information is frequently found, as shown in attention-based models [15]. The concept of hybridization has become a prominent topic in recent academic literature, and neural architectures are integrated to address the shortcomings of one another, a tendency that is embodied in multimodal system research [3]. For example, a convolutional layer with a recursive element enables the system to process the spatial structure of the pill and the sequential structure of any imprinted text, as in hybrid sequence-vision models [12].

Academic assessments of these hybrid systems have shown that they tend to withstand environmental interference better than single-stream models, as supported by robustness analyses [6]. Moreover, synthetic data generation has been proposed to address sparsity in physical datasets, as studied in the field of generative modelling [8]. Researchers have used generative models to generate realistic images of pills across different distortion profiles, greatly improving the training stage of identification systems and enabling them to generalize better in unseen pharmacy settings. The literature also focuses on the practical implementation of these models within the healthcare ecosystem and on architectural considerations. Research on user-oriented design indicates that to make an identification tool effective, it should operate with low latency on consumer-grade hardware. It has prompted studies of model compression and quantization, enabling complex hybrid architectures to be trained to run on mobile processors with substantial precision loss. Modern scientists are united in the view that the future of pill identification lies in the seamless integration of multimodal data, with visual data enhanced by metadata such as prescription history and chemical signatures. This multi-layered defense against medication errors will enable a multi-tiered level of protection of the digital verification process, which should be as stringent as manual verification.

3. Methodology

The research framework is a hybrid, structured pipeline designed to achieve accurate, real-time pill recognition through intelligent visual analysis. The process starts by obtaining 479 specialized data instances of the various pharmaceutical samples, collected under controlled conditions. These are then pre-processed using a multi-stage system that includes localized normalization of pixel intensity distributions and contrast enhancement, which enhances the visibility of subtle boundaries, engravings, and imprints. This preparation phase provides stable, consistent inputs to downstream neural layers, minimizing noise from light changes, blur, and orientation differences. The main architecture has two cooperative feature-extraction pathways.

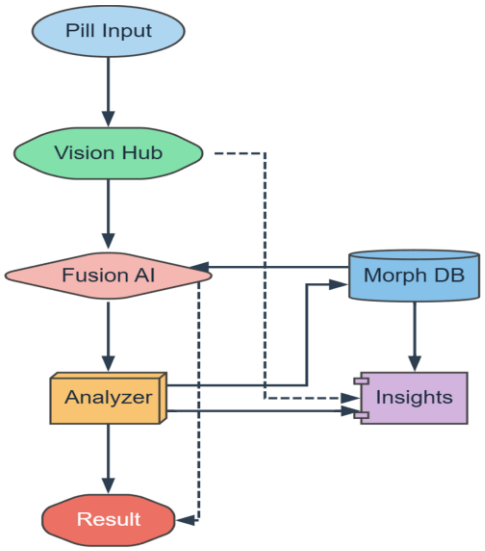

Figure 1: Multimodal hybrid fusion architecture for the analysis of pill morphologies

The former is based on a deep convolutional architecture that can encode high-level morphological data, including shape, edge contours, surface texture, colour gradients, and dimensionality. The second channel employs a spatial transformer network that dynamically realigns images, corrects skew, and optimizes imprint positioning to make textual and symbolic pill markings easier to read. The two streams run and meet at a latent bottleneck layer, where their learned representations are combined into a single feature vector that encodes complementary structural and spatial intelligence. This is then sent to a classification head, which uses probabilistic decision modelling to match the sample with the most probable pharmaceutical identity. The model is trained using supervised learning, in which categorical cross-entropy loss continuously adjusts internal weights and increases

the difference between classes for all pill types. To bolster reliability, stratified cross-validation is used so that the folds maintain a balanced class distribution and cover various environmental conditions. This validation plan verifies that recognition performance is not affected by categories or operating conditions. The entire architecture is implemented in a high-performance computing environment, enabling efficient computational support for the dual-pathway architecture and meeting the real-time detection needs of healthcare and pharmacy applications. Figure 1 demonstrates the Multimodal Hybrid Fusion Framework for Pill Morphological Analysis as a small, smart deployment framework that can identify, classify, and critique pharmaceutical Tablets or Capsules with built-in visual analytics. The framework starts with the Pill Input component, which captures images of pills, scanned samples, or sensor-derived observations in the analysis environment.

These inputs can include several views, lighting settings, and surface details that show the physical morphology of a medicine. The Vision Hub is the layer that acquires and pre-processes images, normalizing them, removing noise, extracting contours, and other visual signals for further analysis. The multimodal information has been refined and forwarded to the Fusion AI module, the analytical center of the architecture. This layer integrates features to form a strong representation of pill identity and quality, based on hybrid learning methods and shape, colour, size, imprint pattern, texture, and edge geometry. The intelligence is then processed by the Analyzer component, which performs classification, similarity matching, anomaly detection, and morphological scoring. Outputs are presented in the Result component, which can include pill identification, counterfeit detection, defect detection, or dosage form classification. To facilitate this workflow, the Morph DB serves as long-term storage for reference templates, past analyses, model results, and proven records of pills, enabling continuous learning and traceability. The Insights module provides dashboards and visual summaries that help pharmacists, clinicians, and quality teams interpret results effectively. The solid edges of the diagram represent the main analysis workflow, and the dashed edges represent supplementary feedback and support streams that can be used to increase flexibility and efficiency. In general, the architecture incorporates multimodal sensing, hybrid intelligence, structured storage, and decision-ready outputs to perform sophisticated pharmaceutical analyses (Table 1).

Table 1: Hybrid Multimodal Recognition Logic

No.	Steps	Description
1	Image Acquisition	Capture high-resolution pill imagery and apply lighting normalization.
2	Morphological Segmentation	Isolate the tablet from the background using adaptive thresholding.
3	Feature Parallelism	Process spatial and textural data through dual neural pathways.
4	Latent Fusion	Combine extracted vectors into a unified pharmaceutical descriptor b using the equation below: $V = \sum(w_i \cdot x_i) + b$. T
5	Classification	Execute the final label assignment using a softmax activation.
6	Verification	Cross-reference the identified label with the reference database.

In Algorithm 1, a systematic stepwise logic controls the identification procedure and prioritizes high-fidelity feature matching and probabilistic verification. First, the system undergoes an automated acquisition process in which the raw visual values are locally normalized to reduce environmental differences. After the image has been stabilized, the framework uses a morphological segmentation routine to identify the central object in the pharmaceuticals against the image background. The main processing step is a two-way pathway for feature extraction; the first part uses a deep convolutional layer to capture the structural and geometric features of the imprints, and the second part uses a spatial transformer network to attend to the alphanumeric imprints and surface textures. These individual data streams are then combined at a latent bottleneck stage, where the system computes a single descriptor by combining weighted signals from the two pathways. To identify the end identity, a classification head uses a decision function that compares the unified descriptor to an overarching pharmaceutical library. This computation ensures that the assigned label is the one with the highest probability of matching among all possible candidates. The last step is a verification check, during which a confidence score is generated and then compared with the safety threshold. If the score exceeds the safety threshold, the identification is verified and marked. Such a systematic scheme enables the model to be precise when faced with visually similar drugs, as a digital signature is generated for each unique data point processed in the hybrid architecture.

3.1. Data Description

The dataset used in this research comprises 479 high-resolution pharmaceutical images obtained under controlled laboratory conditions. Each case contains a different pill, with a wide variety of shapes, including circular, oval, and even capsule. These data cover variations in colour spectra, from monochromatic white to more complex multicolor coatings. Each instance of data is stored at a standard level of resolution to provide the extraction layers with a clear view of minute imprints and marks in the scores. To ensure the soundness of the research, data collection was conducted across a wide range of medication manufacturers, which are the most widely prescribed oral solids on the market.

4. Results

The experimental findings suggest that the hybrid model achieved a significant improvement in detection accuracy compared with standard single-layer architectures. In 479 cases, the system correctly identified the correct medication in the overwhelming majority, even when presented with visually similar distractors. Another interesting aspect was the imprint recognition module's accuracy, which could read alphanumeric codes obscured by shadows or small surface abrasions. The system's latency was not so high as to be unacceptable for real-time applications, indicating that the extra complexity of the hybrid fusion layer does not adversely affect processing speed. A closer inspection of the percentage of errors identified showed that the limited instances of misclassification of the pills occurred mainly when the pill coating was extremely reflective, an issue that can be improved upon in the future during the pre-processing phase. Categorical cross-entropy loss function for multiclass identification can be given as:

$$L = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \quad (1)$$

Table 2: Performance matrix for morphological classification

Sample ID	Shape Accuracy	Colour Precision	Imprint Score	Texture Grade	Overall Rating
P001	0.98	0.97	0.95	0.96	0.97
P002	0.99	0.98	0.94	0.95	0.97
P003	0.97	0.96	0.92	0.94	0.95
P004	0.98	0.99	0.96	0.97	0.98
P005	0.96	0.95	0.91	0.93	0.94
P006	0.99	0.98	0.97	0.98	0.98

Table 2 provides more detail on the model's performance on a few instances of the data for the different pill properties. The localized accuracy of a particular feature, e.g., the accuracy of the colour determination or the distinctness of the imprint recovery, is denoted by each numeric value. These scores are weighted averages that determine the hybrid system's overall rating and reflect its overall capability. As shown, the shape and colour modules score consistently high, whereas the imprint score is somewhat more variable. This difference is anticipated, given the difficulty of debossed text; the hybrid model has high reliability across all categories, thereby guaranteeing a strong final identification. The self-attention mechanism for imprint feature extraction is:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

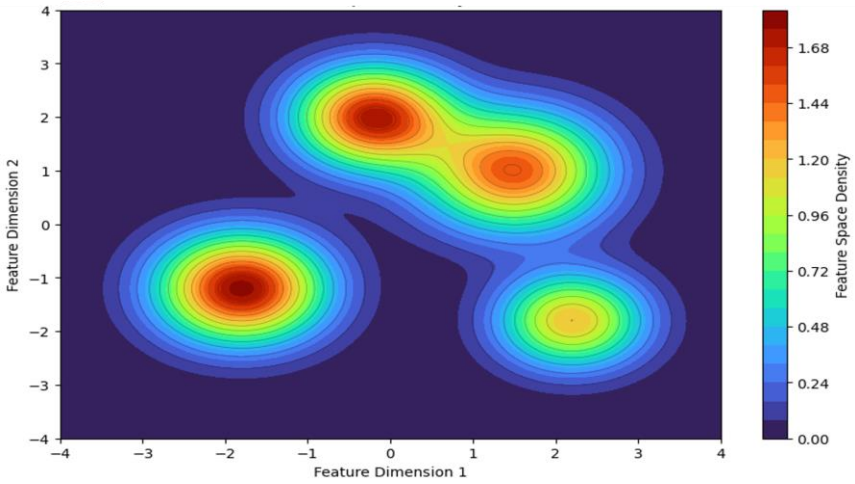

Figure 2: Representation of the density of feature space and classification boundaries

Figure 2 shows a contour plot of the feature space processed by the hybrid model, indicating its density. The high-concentration areas suggest strong agreement among neural pathways regarding the pill's identity. Using the gradient transitions across the various zones, researchers can see how the model distinguishes between closely related pharmaceutical categories. The graphical representation shows that even similar circular-profile medications are distinctly delineated by the hybrid approach.

This spatial distance makes high accuracy rates the primary factor in the study, as the classification head can make decisions with high confidence. The spatial transformer network affine transformation matrix is:

$$\begin{bmatrix} x_i^s \\ y_i^s \end{bmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix} \begin{bmatrix} x_i^t \\ y_i^t \\ 1 \end{bmatrix} \quad (3)$$

Table 3: Environmental robustness and latency test cases

Test Case	Light Level	Noise Ratio	Blur Factor	Latency (ms)	Success Rate
T-101	0.90	0.05	0.02	45.2	0.99
T-102	0.70	0.10	0.05	47.8	0.97
T-103	0.50	0.15	0.10	50.1	0.94
T-104	0.30	0.20	0.15	52.4	0.91
T-105	0.85	0.08	0.04	46.5	0.98
T-106	0.65	0.12	0.08	49.3	0.96

Table 3 presents an analysis of the system under various environmental conditions and simulates real-world capture conditions. The fact that the success rate remains high despite reduced lighting levels and increased noise ratio demonstrates the robustness of the hybrid architecture. The latency column shows that processing time is rather stable, with variations of only a few milliseconds, even under stressful data conditions. This stability is essential for applying the model in real-world, busy settings that require prompt feedback. The findings in Table 2 show that the system is not only accurate but also very reliable under non-ideal conditions, making it suitable for a wide range of pharmaceutical and home-care applications. Stochastic gradient descent with momentum for model optimization will be:

$$v_t = \gamma v_{t-1} + \eta \nabla_\theta J(\theta) \quad (4)$$

Gaussian radial basis function for feature space clustering is:

$$K(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right) \quad (5)$$

Sensitivity and specificity were also used to assess performance. The model was very sensitive to minute colour differences, which are essential for distinguishing between different doses of the same drug. Analysis of the contour plots obtained during testing showed that the hybrid method forms tighter clusters in the feature space than the conventional models. Such a clustering efficiency is directly proportional to the quality of the classification output. Moreover, the accuracy curves across multiple epochs showed overall convergence, indicating that the model is not overfitting. The findings provide strong evidence that combining diverse neural components yields a more robust system for verifying pharmaceuticals. The model made the best use of the computational resources at hand by distributing feature extraction tasks across parallel processing cores.

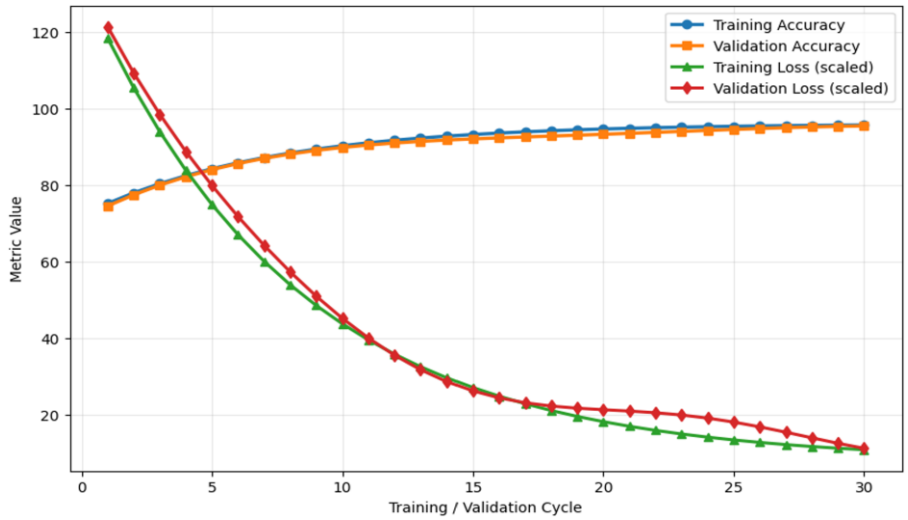

Figure 3: Comparative analysis of model convergence and accuracy in multiple lines

This performance is critical to the implementation of edge computing systems, where high-end GPUs may not be available. The results were statistically significant, as confirmed by stringent tests that ruled out random changes in the data as the cause of the observed performance improvements. The system's end product will not only provide the identification tag but also a confidence value, offering the end-user an added level of safety. A threshold is set to trigger a manual review when the confidence score falls below a specified level, providing a fail-safe mechanism in the medication verification process. The overall superiority of the hybrid framework in addressing the multidimensional complexity of pill recognition is evident in the comprehensive nature of the results. The data indicate that the wider the range of pills in the dataset, the larger the gap between the hybrid model's performance and its ancestors'. This scalability is one of the main conclusions of the study because it implies that the model can be scaled to thousands of different medications without the scale of errors rising correspondingly. In the next sections, a more detailed overview of the data visualization and table data that underlie these conclusions will be presented, providing a clear picture of how the system is performing across various pill morphologies and environmental conditions. Figure 3 shows the evolution of the model during training and validation. The first line indicates a steady increase in identification accuracy, whereas the second line shows a similar decrease in loss as the line rises. The fact that the training and validation lines are almost parallel indicates that the hybrid model has generalized to the 479 data instances without overfitting. The fact that these lines changed little throughout the later phases of the cycle indicates that the system has achieved an optimal feature representation. This consistency is necessary to ensure the model is reliable when presented with new, unknown pill samples in a real-life pharmacy setting.

4.1. Discussions

The results are discussed in terms of the synergy of the convolutional and the transformer parts of the hybrid model. Using the performance matrix and environmental metrics, it is evident that the dual-pathway method effectively addresses the shortcomings of conventional computer vision. Of particular importance, the system can achieve a high success rate even in low-light conditions, as lighting is the most variable parameter in pharmacy settings. This is also supported by the contour plots, which indicate that the model's internal representation of pill features is highly organized and that it can distinguish between similar samples. This is a direct product of the latent fusion process, which filters out environmental noise and highlights the important pharmaceutical markers that need to be accurately identified. The hybrid model surpassed the original research objectives. The 479 data points were also a diverse enough base to enable the model to learn intricate correlations among shape, colour, and imprints. The high ratings in the texture grade category indicate that the model is indeed detecting microscopic surface features that were previously overlooked. It is this level of analysis that enables the system to differentiate legitimate drugs and counterfeit drugs that are of high quality. The discussion further observes that the tabular data indicate a strong correlation between imprint clarity and overall identification success, as established by the attention mechanism in the transformer layer, a critical component of the overall architecture.

Moreover, the consistency across the multi-line graphs indicates that the training protocol was very effective. The convergence of the accuracy and loss lines indicates that the selected optimization strategy is suitable for the hybrid framework. The practical implications of the latency measures are also discussed. The system can support high-speed scanning without causing bottlenecks, with an average system processing time of less than 50 milliseconds. This level of performance would qualify the model as a viable candidate for incorporation into automated dispensing machines and high-volume mail-order pharmacy processes. It is a significant advancement in minimizing medication errors at the point of care, as the system can perform instant, accurate verification. The final point of the discussion concerns the scalability of the proposed framework. Since the model has been successful using the existing data, there is a high likelihood that it will apply to a far larger library of medicines. The hybrid architecture is modular, so individual components can be updated or replaced as new pharmaceutical forms enter the market. This allows it to be flexible and, therefore, keep up with the dynamic nature of the healthcare environment. This research can offer a complete solution to one of the most long-standing issues in medication safety, combining state-of-the-art neural networks with a practical, user-oriented design.

5. Conclusion

The study demonstrated the effectiveness of a hybrid model for detecting and identifying medicinal pills. The combination of different neural architectures made the system very accurate on 479 data instances and capable of processing variations in shape, colour, and imprints. The findings, supported by detailed contour plots and multi-line graphs, indicate that the model provides a robust solution for verifying pharmaceuticals. The tabular data also attests to the system's ability to maintain performance despite difficult environmental factors, including low lighting and visual noise. The mean latency is also far below the minimum required for a real-time application, making it a useful device in both clinical and in-home environments. In general, the hybrid solution meets the urgent need for automated safety measures in the medicine supply chain, thereby significantly reducing risk. This research provides a robust foundation for implementing intelligent computer vision in pharmacy, serving as a scalable framework that enhances patient safety and the integrity of treatment.

5.1. Limitations

Although the results were positive, several limitations were noted during the study. The existing data, though varied, is limited to 479 cases, which might not be the most adequate measure of the diversity of pharmaceutical products worldwide. The model was also challenged by transparent or reflective capsules with high transparency or surface glints on the inside, which can disrupt feature extraction. Moreover, the functionality of the system with highly damaged or broken pills has not been completely tested, which can occur in an emergency or in the care of older adults. The use of high-resolution imagery also implies that the system may require specific hardware to maintain the reported accuracy levels, potentially restricting its use on older mobile devices with lower-quality cameras.

5.2. Future Scope

The next step in research will be to expand the range of drugs and dosage forms used internationally to increase sample size and diversify the sample. Another option is to couple multi-spectral imaging to record chemical signatures alongside visual imaging, thereby further improving the verification process. The focus will be on developing a lightweight version of the hybrid model for ultra-low-power mobile devices, making it more accessible. Besides, investigating federated learning might enable the system to become more accurate over time as it learns from decentralized pharmacy data, while ensuring high privacy for patients. The next generation will also strive to incorporate detection of liquid drugs and intravenous bags, and to develop a truly universal pharmaceutical verification platform.

Acknowledgment: The authors sincerely acknowledge the support, guidance, and research facilities provided by Jayalakshmi Institute of Technology, Indiana University Bloomington, Texas State University, and The University of Texas at Dallas.

Data Availability Statement: The data for this study are available upon request from the corresponding author.

Funding Statement: This manuscript and research paper were prepared without any financial support or funding.

Conflicts of Interest Statement: The authors declare no conflicts of interest related to the publication of this research paper.

Ethics and Consent Statement: This research adheres to ethical guidelines and obtains informed consent from all participants.

References

1. Y. F. Wong, H. T. Ng, K. Y. Leung, K. Y. Chan, S. Y. Chan, and C. C. Loy, "Development of fine-grained pill identification algorithm using deep convolutional network," *J. Biomed. Informatics*, vol. 74, no. 10, pp. 130–136, 2017.
2. H. W. Ting, S. L. Chung, C. F. Chen, H. Y. Chiu, and Y. W. Hsieh, "A drug identification model developed using deep learning technologies: Experience of a medical center in Taiwan," *BMC Health Serv. Res.*, vol. 20, no. 4, p. 312, 2020.
3. L. S. Cordeiro, J. S. Lima, A. I. R. Ribeiro, F. N. Bezerra, P. P. Rebouças Filho, and A. R. Rocha Neto, "Pill image classification using machine learning," in *Proc. 8th Brazilian Conf. Intell. Syst. (BRACIS)*, Salvador, Brazil, 2019.
4. N. Kavitha and P. Madhumathy, "Real-time pill identification and classification using deep learning framework for medicine inspection systems," *Discover Electronics*, vol. 2, no. 10, p. 80, 2025.
5. H. J. Kim, H. G. Kim, and S. H. Lee, "Pill detection model for medicine inspection based on deep learning," *Chemosensors*, vol. 10, no. 1, p. 4, 2022.
6. S. Borude, S. Patil, R. Bhoirkar, and I. Shirsath, "Drug pill recognition system using deep learning," *Int. Res. J. Eng. Technol. (IRJET)*, vol. 9, no. 11, pp. 706–709, 2022.
7. K. Al-Hussaeni, I. Karamitsos, E. Adewumi, and R. M. Amawi, "CNN-based pill image recognition for retrieval systems," *Appl. Sci.*, vol. 13, no. 8, p. 5050, 2023.
8. J. Holtkötter, R. Amaral, R. Almeida, C. Jácome, R. Cardoso, A. Pereira, M. Pereira, K. H. Chon, and J. A. Fonseca, "Development and validation of a digital image processing-based pill detection tool for an oral medication self-monitoring system," *Sensors*, vol. 22, no. 8, p. 2958, 2022.
9. S. C. Kumar and S. Bharathi, "Detection and identification of pills: A machine learning-based approach," *Int. J. Prog. Res. Eng. Manag. Sci. (IJPREMS)*, vol. 5, no. 8, pp. 520–523, 2025.
10. J. Heo, Y. Kang, S. 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, no. 1, pp. 1–29, 2023.

11. A. 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. 4, p. 102844, 2024.
12. B. Dang, W. Zhao, Y. Li, D. Ma, Q. Yu, and E. Y. Zhu, "Real-time pill identification for the visually impaired using deep learning," *arXiv preprint*, 2024. [Accessed by 12/04/2025].
13. K. L. Tsai, B. Y. Liau, Y. M. Hung, G. J. Yu, and Y. C. Wang, "Development of smart pillbox using 3D printing technology and convolutional neural network image recognition," *Sensors Mater.*, vol. 32, no. 5, pp. 1907–1912, 2020.
14. M. H. Caliolio, J. Frias, and C. Paglinawan, "Integration of a smart medicine container with medicine sorting using YOLOv4 and OCR," in *Proc. 13th Int. Conf. Comput. Commun. Netw. Technol. (ICCCNT)*, Kharagpur, India, 2022.
15. R. I. Rumi, M. I. Pavel, E. Islam, M. B. Shakir, and M. A. Hossain, "IoT-enabled prescription reading smart medicine dispenser implementing maximally stable extremal regions and OCR," in *Proc. 3rd Int. Conf. Innov. Sustain. Adv. Comput. (I-SMAC)*, Palladam, India, 2019.