



Tạp chí điện tử
Khoa học và Công nghệ Trường Đại học Công nghệ Đông Á
Website Tạp chí: <https://vjai.org.vn>

Application of Computer Vision for Pallet Stacking in Automated Warehouse Operations

Phung Van Thom
Faculty of Mechanical Engineering, East Asia University of Technology

ARTICLE INFO

Keywords:

pallet detection,
warehouse automation,
material handling,
image processing

ABSTRACT

This study addresses the problem of accurately detecting and quantifying the misalignment between two stacked pallets, aiming to automate the loading and unloading process in warehouse environments. The authors propose an integrated approach that employs an RGB camera in combination with the YOLOv11 deep learning network for pallet detection. Subsequently, the Hough transform is applied to extract boundary lines and key points of the pallets. To determine the alignment condition, the Euclidean distance between corresponding edges of the target pallets is calculated. Experimental results demonstrate the system's capability to accurately recognize pallets within the forklift's field of view. This enables the forklift to promptly adjust its position to place goods correctly, thereby preventing the risk of load collapse when stacking at high levels. The proposed method not only provides an effective solution for automating logistics operations but also offers substantial potential for application, particularly in meeting the demands of automated operation and intelligent management in large-scale warehouse systems.

* Corresponding author

Email: thompv@eaut.edu.vn

DOI: <https://doi.org/10.65153/3gf5f910>

Received: 19/06/2025; Received in revised form: 27/03/2026; Accepted: 30/03/2026

Available online: 01/04/2026

Published by: East Asia University of Technology



Ứng dụng thị giác máy tính hỗ trợ xếp chồng pallet trong vận hành kho hàng tự động

Phùng Văn Thơm
Khoa Cơ khí, Trường Đại học Công nghệ Đông Á

THÔNG TIN BÀI BÁO

TÓM TẮT

Từ khóa:

nhận dạng pallet,
tự động hoá kho bãi,
xếp dỡ hàng hoá,
xử lý hình ảnh

Nghiên cứu này giải quyết bài toán phát hiện và định lượng chính xác độ lệch giữa hai pallet xếp chồng nhằm mục tiêu tự động hóa quá trình xếp dỡ hàng hóa trong môi trường kho bãi. Tác giả đề xuất một phương pháp tích hợp sử dụng camera RGB kết hợp mạng học sâu YOLOv11 cho nhiệm vụ phát hiện pallet. Sau đó, biến đổi Hough để trích xuất các đường ranh giới và điểm chính của pallet. Để xác định tình trạng xếp thẳng hàng hóa, phương pháp tính khoảng cách Euclid giữa hai cạnh cùng phía của pallet mục tiêu đã được áp dụng. Kết quả thực nghiệm minh chứng khả năng nhận dạng pallet chính xác của hệ thống trong trường quan sát của xe nâng. Điều này cho phép xe nâng điều chỉnh tư thế kịp thời để đưa hàng vào đúng vị trí, qua đó ngăn ngừa rủi ro đổ hàng hóa khi xếp chồng cao. Phương pháp này không chỉ cung cấp một giải pháp hiệu quả cho việc tự động hóa các quy trình hậu cần mà còn mở ra tiềm năng ứng dụng lớn, đặc biệt đáp ứng nhu cầu vận hành tự động và quản lý thông minh của các kho bãi có quy mô rộng lớn.

1. MỞ ĐẦU

Trong bối cảnh bùng nổ của cuộc cách mạng công nghiệp 4.0, tự động hóa kho hàng đã trở thành một yếu tố then chốt, mang lại tiềm năng to lớn trong việc nâng cao hiệu suất, giảm thiểu chi phí vận hành và tối ưu hóa toàn bộ chuỗi cung ứng [1]. Tuy nhiên, một trong những thách thức hàng đầu trong quá trình chuyển đổi này là việc thực hiện các tác vụ xếp dỡ và xếp chồng pallet một cách chính xác, đặc biệt khi yêu cầu các chồng hàng có độ cao lớn. Thực tế cho thấy, ngay cả những sai lệch vị trí nhỏ nhất giữa các pallet cũng có thể dẫn đến hậu quả nghiêm

trọng như đổ vỡ hàng hóa, gây thiệt hại đáng kể về kinh tế và tạo ra rủi ro mất an toàn lao động. Các phương pháp truyền thống, dù là dựa vào sự can thiệp thủ công hay sử dụng các hệ thống cảm biến đơn giản, thường bộc lộ nhiều hạn chế về độ tin cậy và hiệu quả, đặc biệt trong môi trường vận hành đòi hỏi tốc độ cao và tự động hóa hoàn toàn của các kho hàng hiện đại. Để giải quyết triệt để bài toán này, thị giác máy tính nổi lên như một công nghệ đột phá, cung cấp một giải pháp đầy hứa hẹn. Khả năng vượt trội của thị giác máy tính trong việc nhận diện, phân tích và diễn giải hình ảnh mở ra một hướng tiếp cận hoàn toàn

* Tác giả liên hệ

Email: thompv@eaut.edu.vn

DOI: <https://doi.org/10.65153/3gf5f910>

Ngày nhận: 19/06/2025; Ngày nhận bản sửa: 27/03/2026; Ngày chấp nhận: 30/03/2026

Ngày online: 01/04/2026

Đơn vị xuất bản: Trường Đại học Công nghệ Đông Á

mới để giám sát và điều khiển các tác vụ xếp chồng pallet một cách tự động và đạt độ chính xác cao. Việc tích hợp công nghệ này vào các hệ thống xe nâng tự động hoặc robot công nghiệp có thể cho phép phát hiện và định lượng tức thời độ lệch giữa các pallet. Từ đó, hệ thống có thể đưa ra các tín hiệu điều chỉnh kịp thời, đảm bảo rằng quá trình xếp chồng diễn ra một cách an toàn, ổn định và hiệu quả tối đa.

Nghiên cứu này tập trung vào việc ứng dụng thị giác máy tính để hỗ trợ xếp chồng pallet trong vận hành kho hàng tự động. Tác giả đề xuất một phương pháp tích hợp tiên tiến, sử dụng camera RGB kết hợp với mạng học sâu YOLOv11 [2, 3] cho nhiệm vụ phát hiện pallet chính xác. Tiếp theo, biến đổi Hough [4] sẽ được áp dụng để trích xuất các đường ranh giới và điểm chính của pallet. Để xác định tình trạng xếp thẳng hàng, nghiên cứu sử dụng phương pháp tính khoảng cách Euclid [5] giữa hai cạnh cùng phía của pallet mục tiêu. Kết quả thực nghiệm dự kiến sẽ minh chứng khả năng nhận dạng pallet chính xác của hệ thống trong trường quan sát của xe nâng. Điều này không chỉ giúp xe nâng tự động điều chỉnh tư thế kịp thời để đưa hàng vào đúng vị trí, mà còn đóng vai trò then chốt trong việc ngăn ngừa rủi ro đổ hàng hóa khi xếp chồng cao. Phương pháp được đề xuất không chỉ cung cấp một giải pháp hiệu quả cho việc tự động hóa các quy trình hậu cần mà còn mở ra tiềm năng ứng dụng rộng lớn, đặc biệt đáp ứng nhu cầu vận hành tự động và quản lý thông minh của các kho bãi có quy mô lớn trong tương lai.

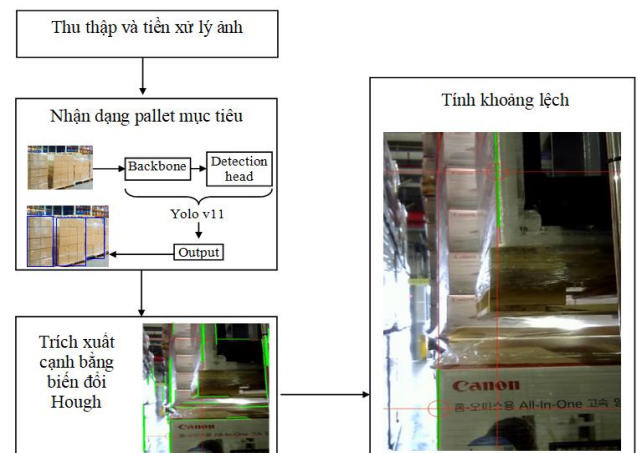
2. GIẢI QUYẾT VẤN ĐỀ

Sơ đồ luồng của phương pháp đề xuất được minh họa trong Hình 1, bao gồm bốn giai đoạn chính: (1) thu thập và xử lý hình ảnh, (2) nhận dạng pallet mục tiêu và xác định cạnh, (3) trích xuất điểm chính và (4) tính toán khoảng lệch. Đầu tiên, hình ảnh RGB của pallet được chụp bởi camera RGB đặt tại vị trí quan sát của xe nâng. Mạng học sâu YOLOv11 được ứng dụng để nhận dạng pallet từ hình ảnh 2D. Tiếp theo, các đường thẳng ranh giới của pallet được trích xuất bằng Biến đổi Hough. Trung điểm của các cạnh pallet sau đó được sử dụng để xác định khoảng cách lệch giữa hai chồng hàng.

2.1. Thu thập và tiền xử lý hình ảnh

Sau khi các kệ hàng pallet được xếp trên mặt đất, xe nâng không người lái sẽ nhận nhiệm vụ xúc kệ pallet mục tiêu và di chuyển đến vị trí

tương ứng được thiết lập sẵn tại kho để xếp chồng lên pallet khác.



Hình 1. Tổng quan phương pháp xác định độ lệch pallet

Ban đầu, xe nâng cần đứng tại vị trí quan sát được thể hiện rõ trong Hình 2, trong đó D biểu thị khoảng cách theo chiều dọc giữa vị trí quan sát và đường dừng. Việc xác định giá trị D cần tuân thủ các điều kiện để đảm bảo an toàn: (i) Pallet phải luôn nằm trong trường nhìn của camera, ngay cả khi xe nâng hoặc pallet được xếp lệch; (ii) Khoảng cách phạm vi quan sát: Vị trí quan sát không được quá xa vạch dừng, bởi độ chính xác của camera giảm đáng kể khi khoảng cách tăng. Đồng thời, khoảng cách này cũng không được quá gần pallet, tránh nguy cơ mất thông tin quan trọng gây ảnh hưởng đến quá trình nhận dạng pallet.



Hình 2. Minh họa xe nâng sử dụng để bốc dỡ hàng hóa

Xe nâng không người lái được trang bị camera RGB để thu thập hình ảnh của pallet liên tục trong thời gian hoạt động. Tuy nhiên, khi trong môi trường làm việc, một số hiện tượng ánh sáng không đồng đều có thể xảy ra, để giảm thiểu

hiều và tập trung vào đối tượng quan tâm, các kỹ thuật tiền xử lý hình ảnh được áp dụng. Một trong những phương pháp phổ biến là sử dụng các bộ lọc miền không gian tuyến tính. Cụ thể, bộ lọc Gaussian [6] sử dụng hàm Gaussian 2D để tính trung bình có trọng số các pixel lân cận giúp giảm thiểu hiện tượng răng cưa của ảnh:

$$G(u, v) = \frac{1}{2\pi\sigma^2} e^{-\frac{u^2+v^2}{2\sigma^2}} \quad (1)$$

Trong đó u, v là tọa độ tương đối từ tâm của hạt nhân, σ là độ lệch chuẩn điều chỉnh mức độ làm mờ ảnh.

Do áp dụng biến đổi Hough nên các cạnh cần có mức độ nổi bật, nghiên cứu sử dụng bộ lọc làm sắc nét ảnh bằng toán tử Laplacian [7] bằng sử dụng phép tích chập với các hạt nhân có kích thước 3×3 mô tả bởi:

$$(I * K)(x, y) = \sum_u \sum_v I(x-u, y-v) \cdot K(u, v) \quad (2)$$

Trong đó $I(x, y)$ là giá trị cường độ pixel tại vị trí (x, y) trong ảnh đầu vào, $K(u, v)$ là giá trị của phần tử tại vị trí (u, v) trong ma trận hạt nhân Laplacian với hai ma trận được xét đến tương ứng là các lặn cận trực tiếp và các lặn cận chéo mô tả như sau:

$$\mathbf{K}_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}; \quad \mathbf{K}_2 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Cuối cùng, mức độ độ sắc nét của ảnh $I'(x, y)$ thu được bằng cách điều chỉnh hệ số dương k với toán tử Laplacian được cho bởi:

$$I'(x, y) = I(x, y) + k \cdot \text{Laplacian}(I(x, y)) \quad (3)$$

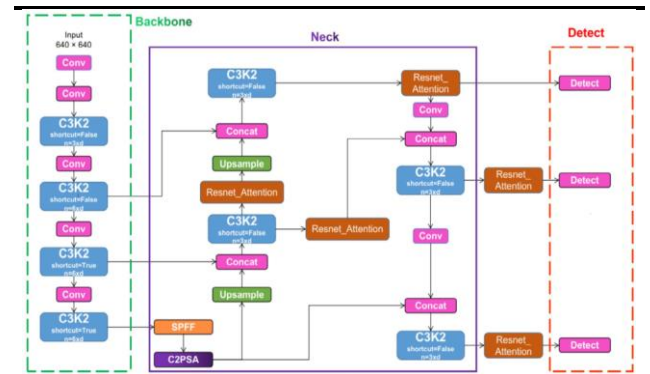
2.2. Nhận dạng pallet mục tiêu

Trong các kho hàng tự động có nhiều loại pallet khác nhau, sự lưu chuyển liên tục của pallet giữa các kho bãi dẫn đến mức độ hao mòn và biến dạng đáng kể của chúng và hàng hoá được xếp trên nó. Hàng hoá có thể bị nghiêng ở góc độ nhỏ, bị che khuất một phần bởi màng bọc nhựa hoặc chịu ảnh hưởng từ các điều kiện môi trường khác nhau. Trong những tình huống này, các phương pháp xử lý ảnh truyền thống gặp khó khăn trong việc xử lý hiệu quả. Ngược lại, các phương pháp phát hiện đối tượng dựa trên học sâu, thông qua việc trích xuất các đặc trưng qua mạng nơ-ron đa lớp có khả năng phân loại và

định vị mục tiêu, do đó phù hợp hơn cho các tình huống nêu trên. Đối với xe nâng không người lái, việc tối ưu hóa thời gian suy luận để đạt được kết quả phát hiện nhanh chóng là rất quan trọng. Do đó, nghiên cứu áp dụng mô hình YOLOv11 cho nhiệm vụ phát hiện pallet. YOLOv11 nổi bật với khả năng phát hiện đối tượng theo thời gian thực với độ chính xác cao và hiệu quả tính toán, đạt được sự cân bằng tốt giữa độ chính xác và tốc độ xử lý. Bảng 1 sẽ mô tả kiến trúc chính của YOLOv11 được sử dụng.

Bảng 1. Các lớp chính trong kiến trúc phát hiện vật thể của YOLO v11

Lớp	Vai trò
Convolution + Batch Normalization + Sigmoid Linear Unit	Lớp cơ bản xử lý đặc trưng
C3 block	Khối tích chập chồng học đặc trưng trừu tượng
SPPF (Fast pooling)	Gộp thông tin
Upsample	Mở rộng kích thước đặc trưng để kết nối giữa tầng cao và thấp
Detect Layer	Trả về khung biên dạng + lớp + vị trí



Hình 3. Cấu trúc chi tiết của YOLO v11 [8]

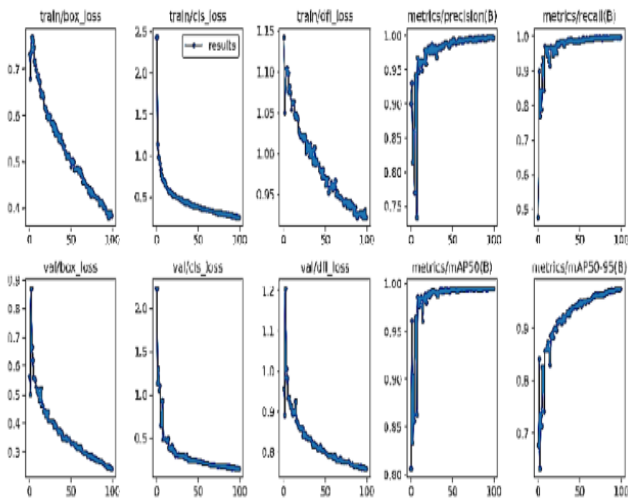
Để đảm bảo khả năng ứng dụng trong môi trường thực tế, dữ liệu huấn luyện được thu thập bao gồm các pallet và hàng hoá xếp chồng trên đó trong nhiều điều kiện khác nhau. Ngoài ra, bộ dữ liệu còn bao gồm các pallet bị che khuất một phần bởi màng bọc nhựa ở các độ cao khác nhau hoặc ở các góc nghiêng nhỏ, như được minh họa trong Hình 4.

đáng tin cậy. Một số kết quả nhận dạng được trình bày tại Hình 6.



Hình 4. Một số hình ảnh các pallet mục tiêu trong tập huấn luyện dưới các điều kiện khác nhau ở một số góc nhìn khác nhau, có hàng hoá đã xếp sẵn trên pallet

Bộ dữ liệu cho nhiệm vụ phát hiện pallet bao gồm 1275 ảnh RGB. Trong đó, 80% dữ liệu được dành để huấn luyện, trong khi 20% được sử dụng cho tập xác thực mà không thiết lập tập kiểm tra riêng biệt. Đối với các ảnh RGB có độ phân giải 640×480 , kích thước lô được đặt là 32. Tốc độ học ban đầu được đặt là 0,0003 và được điều chỉnh động trong 100 epoch huấn luyện.



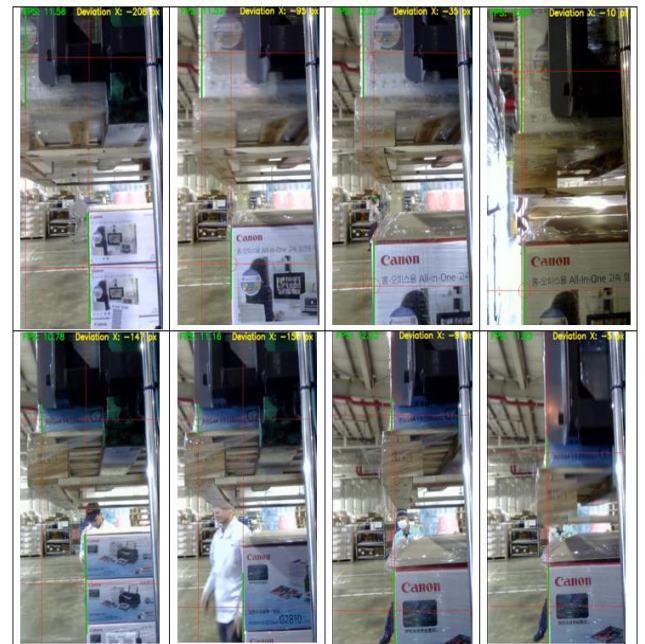
Hình 5. Kết quả theo dõi quá trình huấn luyện qua từng vòng

Từ kết quả Hình 5 cho thấy khả năng học và tổng quát hóa tốt của mô hình tương đối tốt. Các đường cong mất mát trên cả tập huấn luyện (train/box_loss, train/cls_loss, train/obj_loss) và tập xác thực (val/box_loss, val/cls_loss, val/obj_loss) đều giảm đều và ổn định qua 100 epoch. Đồng thời, các chỉ số hiệu suất như precision, recall, mAP50 và mAP50-95 đều tăng nhanh chóng và đạt mức cao ổn định sau khoảng 20-30 epoch. Điều này chứng tỏ mô hình có khả năng định vị và phân loại đối tượng chính xác và



Hình 6. Kết quả kiểm thử trên tập dữ liệu xác thực

2.3. Trích xuất cạnh sử dụng biến đổi Hough và tính độ lệch



Hình 7. Kết quả đo độ lệch thực tế khi vận hành

Sau khi pallet được YOLOv11 phát hiện, các đường thẳng ranh giới của chúng sẽ được trích xuất bằng Biến đổi Hough. Biến đổi Hough [9] là một kỹ thuật mạnh mẽ để phát hiện các hình dạng nhất định (đường thẳng, hình tròn) trong ảnh bằng cách ánh xạ các điểm ảnh từ không gian Descartes sang không gian tham số, biểu diễn dưới dạng cực cho bởi:

$$\rho = x \cos \theta + y \sin \theta \quad (4)$$

Trong đó ρ là khoảng cách vuông góc từ gốc tọa độ đến đường thẳng và θ là góc của pháp tuyến từ gốc tọa độ đến đường thẳng. Mỗi điểm ảnh (x, y) trên đường thẳng sẽ tạo thành một đường cong hình sin trong không gian tham số (ρ, θ) , và

sự giao nhau của nhiều đường cong này cho phép xác định các đường thẳng chính xác trong ảnh. Để đảm bảo tính linh hoạt, đặc biệt khi pallet có thể không được đặt hoàn toàn nằm ngang, dải ngưỡng góc ngang được đặt là $[-3^\circ; 3^\circ]$ và dải ngưỡng góc dọc là $[-52^\circ; 52^\circ]$. Từ ảnh RGB, chỉ những đoạn thẳng ngang dài hơn 120 pixel và đoạn thẳng dọc dài hơn 20 pixel được trích xuất.

Sau khi các đường ranh giới của pallet đã được trích xuất bằng Biến đổi Hough, nghiên cứu tiến hành xác định các điểm chính của pallet. Cụ thể, trung điểm của các cạnh pallet sẽ được tính toán. Những trung điểm này sau đó được sử dụng để xác định khoảng cách lệch giữa hai chõng hàng bằng khoảng cách Euclid như sau:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (5)$$

Trong đó d là độ lệch giữa 2 chõng hàng (đơn vị pixel), (x_1, y_1) là tọa độ trung điểm cạnh thứ nhất và (x_2, y_2) là tọa độ trung điểm cạnh thứ hai. Thông tin này cung cấp thông tin định lượng về độ lệch giữa các chõng hàng liên tiếp cho hệ thống điều khiển xe nâng. Một số kết quả tính toán độ lệch được mô tả chi tiết trong Hình 7.

3. KẾT LUẬN

Nghiên cứu này đã thành công trong việc xây dựng một hệ thống thị giác máy tính hỗ trợ xếp chõng pallet, giải quyết các thách thức từ môi trường kho hàng thực tế. Sự kết hợp giữa khả năng phát hiện đối tượng mạnh mẽ của YOLOv11, kỹ thuật trích xuất cạnh chính xác của Biến đổi Hough và phương pháp tính toán độ lệch bằng công thức Euclid đã tạo nên một giải pháp đáng tin cậy. Kết quả đạt được cho thấy tiềm năng lớn trong việc nâng cao hiệu quả và tự động hóa hoạt động xếp dỡ pallet, góp phần cải thiện năng suất và an toàn trong các hệ thống kho hàng tự động hiện đại. Hướng phát triển tiếp theo có thể bao gồm việc thử nghiệm trong môi trường thực tế phức tạp hơn và tích hợp trực tiếp với hệ thống điều khiển của xe nâng tự động.

TÀI LIỆU THAM KHẢO

[1]. Fatorachian, H., & Kazemi, H. (2021). Impact of Industry 4.0 on supply chain performance. *Production Planning & Control*, 32(1), 63-81.

[2]. Khanam, R., & Hussain, M. (2024). Yolov11: An overview of the key architectural enhancements. *arXiv preprint arXiv:2410.17725*.

[3]. Mao, M., & Hong, M. (2025). YOLO Object Detection for Real-Time Fabric Defect Inspection in the Textile Industry: A Review of YOLOv1 to YOLOv11. *Sensors (Basel, Switzerland)*, 25(7), 2270.

[4]. Mukhopadhyay, P., & Chaudhuri, B. B. (2015). A survey of Hough Transform. *Pattern Recognition*, 48(3), 993-1010.

[5]. Dokmanic, I., Parhizkar, R., Ranieri, J., & Vetterli, M. (2015). Euclidean distance matrices: essential theory, algorithms, and applications. *IEEE Signal Processing Magazine*, 32(6), 12-30.

[6]. Ramesh, G., Logeshwaran, J., Gowri, J., & Mathew, A. (2022). The management and reduction of digital noise in video image processing by using transmission based noise elimination scheme. *ICTACT Journal on Image & Video Processing*, 13(1).

[7]. Khan, S. U. (2023). Noisy Image Quality Improvement Using Combinational Filter Models and Sharpening (Master's thesis, Universiti Tun Hussein Onn (Malaysia)).

[8]. Hameed, A., Shah, S. K., Khan, S. U., Alanazi, S., & Algamdi, S. A. (2025). Dermatology 2.0: Deploying YOLOv11 for Accurate and Accessible Skin Disease Detection: A Web-Based Approach. *International Journal of Imaging Systems and Technology*, 35(2), e70050.

[9]. Lapušinskij, A., Suzdalev, I., Goranin, N., Janulevičius, J., Ramanauskaitė, S., & Stankūnavičius, G. (2021). The application of Hough transform and Canny edge detector methods for the visual detection of cumuliiform clouds. *Sensors*, 21(17), 5821.