

Article

Automated Image-Based Detection and Measurement of Wheal Responses in Skin Prick Test Images

Zeenal Punamiya^{1, *}  and Shravani Desai¹ 

¹ Medical Innovation Creativity and Entrepreneurship (M.I.C.E) Lab, Grant Government Medical College and Sir J.J. Group of Hospitals, 6th Floor, New Administration Building, J.J. Hospital Campus, Byculla, Mumbai 400008, India

* Correspondence: drzeenalpunamiya@gmail.com (Z.P.)

Abstract

Allergic disorders are common conditions caused by immune responses to otherwise harmless substances. The Skin Prick Test (SPT) is commonly employed to evaluate sensitivity to particular allergens, with the development of wheals acting as a significant visible reaction. However, manual identification and measurement of wheal size from SPT images may introduce observer variability. This study presents an automated image-analysis system for detecting and quantitatively measuring wheal responses in SPT images. The system was developed using Streamlit, OpenCV, and Imutils for image preprocessing, colour segmentation, contour detection, reference-based calibration, and wheal measurement. On the 30-image held-out evaluation subset, the mean absolute error between the automated measurements and the available reference measurements was 0.48 ± 0.12 mm. The previously reported detection accuracy and false positive/false negative rates could not be maintained as the source dataset lacked sufficient per-wheal ground truth to reconstruct the TP, FP, FN and TN counts independently. The system is computationally lightweight and operates through a standard web interface without specialized hardware. These findings support the feasibility of automated wheal detection and measurement as a computer-assisted approach to SPT image analysis. The system does not independently diagnose allergy or identify the causative allergen and further validation using independently annotated clinical datasets is required.

Keywords: skin prick test; image processing; computer vision; streamlit; wheal detection; automated measurement

Citation: Punamiya, Z., & Desai, S. (2026). Automated Image-Based Detection and Measurement of Wheal Responses in Skin Prick Test Images. *Impact in Computics* 2026, 2, 6. <https://doi.org/10.65500/computics-2026-006>

Received: 17 July 2026 | Revised: 2 September 2026 | Accepted: 3 September 2026 | Published: 9 September 2026

Copyright: © 2026 by the authors. Licensee Impaxon, Malaysia. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Allergic diseases are one of the most common chronic health problems today. They are found in all people across all ages. They are found globally. Allergies develop due to a hyperactive immune system response to particles that are generally harmless. These particles include airborne pollen, dust mites, animal dander, food, and some drugs. Sensitized people usually do not have a troublesome response to these [1], [2]. Identifying the allergens that

cause the most problems to a patient is essential to treat the allergy. This also helps plan the best approach to immunotherapy.

The Skin Prick Test (SPT) is a cost-effective and time-efficient diagnostic tool to evaluate allergen sensitivities. Multiple allergens can be evaluated simultaneously, and test results can be obtained in 15 to 20 minutes. Allergen extracts are applied to the dermis of test subjects and administered via a prick. Positive reactions in subjects are indicated through a raised area of skin, or wheal, at the test

site. Multiple wheals can be evaluated and measured relative to a histamine-positive control [3], [4].

Despite its widespread use, SPT interpretation commonly involves visual assessment and manual measurement by clinical personnel. Differences in wheal delineation and measurement can contribute to variability between observers and across clinical settings [3], [5]. Accurate measurement may also be challenging when appropriate equipment or trained personnel are limited. These limitations necessitate the use of standardized and objective procedures to detect and quantify the wheal responses in the SPT images.

Recent advances in computer vision and machine learning [6], [7] have enabled automatic analysis of medical images, including image segmentation and detection of clinically relevant visual features. Moreover, automated measurement of SPT has been studied in previous works by means of image processing techniques, 3D imaging, specific devices, and deep-learning-based segmentation [8]–[11]. However, the diversity of imaging requirements, computing requirements, and deployment environments still leaves room for more straightforward, more easily accessible image-based approaches.

The proposed system is intended as a computer-assisted tool for SPT image analysis and does not independently diagnose allergy, clinical sensitization, or identify the causative allergen. The proposed system combines classical computer vision methods such as HSV segmentation, Canny edge detection, morphological processing, contour analysis, and reference-based pixel-to-millimetre calibration. A further exploratory Support Vector Machine (SVM) module utilizes predefined computational categories based on wheal diameter rather than independently established clinical severity labels. The system is developed using Streamlit, OpenCV, and Imutils and is designed to run in a standard web browser without requiring specialised hardware. The system was evaluated using images from a publicly available dataset and is intended to automatically detect and measure wheals rather than diagnose allergy, clinical sensitization, or identify the causative allergen.

2. Problem Statement

The Skin Prick Test is a well-established diagnostic method, but manual measurement and visual interpretation suffer from a number of practical disadvantages. Variability in the measurement and interpretation of SPT responses may arise from differences in the way individual clinicians identify and record wheal

sizes, especially if tests are performed at different facilities or by personnel with different training levels. Studies [3], [5] have reported inter-observer variability in SPT measurements, which may affect the reproducibility of measurements in clinical and research settings.

A drawback of conventional SPT analysis is that it is time-consuming and may be observer-dependent. Such manual identification, measurement, and recording of each individual wheal would increase the workload of the clinical staff and could interfere with the assessment process. Decreasing the time required for image-based wheal assessment could improve operational efficiency in primary care and outpatient settings with high patient volumes.

In healthcare settings in low- and middle-income countries, resources may be limited, and standardized allergy assessment may not be possible. Specialized equipment and trained staff are not always available in rural or community health care settings [1]. Advances in computer vision and machine learning have enabled automatic analysis of various medical images. Some of the current methods are based on specialized imaging devices, controlled acquisition parameters, or computational resources that might restrict their practical deployment [9], [10].

The proposed system is intended to assist in standardised wheal assessment rather than serve as an independent diagnostic system. It enables automated image-based wheal detection and measurement through a web browser without requiring specialised imaging hardware.

3. Related work

The computer vision-based medical image analysis and allergy diagnostics have been studied in depth. In this section, we review the most relevant contributions in both fields and previous attempts at automating SPT interpretation.

3.1. Allergy Diagnosis and the Skin Prick Test

European standards for the performance and interpretation of skin prick testing have been described by Heinzerling et al. [3], who also mention that differences in testing procedures, devices, and measurement methods may influence the consistency and interpretation of SPT results.

Serum specific immunoglobulin E (IgE) testing is an alternative diagnostic tool, especially when SPT is not feasible. However, laboratory-based IgE testing is more expensive than SPT and may be less practical for routine first-line screening in resource-limited settings [12]. The

diagnosis of allergic disease is based on SPT together with IgE testing. Both need to be improved in terms of accessibility and objectivity.

3.2. Computer Vision in Dermatological and Medical Imaging

The diagnostic potential of image-based AI was demonstrated in the study by Esteva et al., who trained image-based AI on dermoscopic images and showed that it can classify skin lesions with an accuracy similar to that of board-certified dermatologists [7].

The Hue-Saturation-Value (HSV) color space has been demonstrated to be more robust than the standard Red-Green-Blue (RGB) space for color-based segmentation of skin reactions, since it separates color information from brightness and the influence of uneven illumination is therefore reduced [13]. The Canny edge detector [14], proposed by Canny (1986), remains a popular technique to find object boundaries in images, as it provides an optimal trade-off between high noise filtering and the preservation of real edges [14].

3.3. Automated SPT Analysis

Wöhrle et al. proposed an automatic method for the determination of the wheal size based on image-processing techniques [8]. Pineda et al. [9] proposed an automatic SPT-reading method based on 3D imaging and parametric surface fitting. Justo et al. [10] developed a specific device for the automated reading of prick tests. Attention U-Net model was recently used by Lee et al. for the automatic segmentation of allergy wheals and erythema [11].

Support vector machines (SVMs) are an effective approach to classify medical images based on numerical features extracted from the images and have been successfully used for image characterization [15,16]. On this basis, the proposed system combines classical image-processing techniques, such as color segmentation, contour detection, and feature extraction, with a lightweight SVM classifier in a fully web-accessible interface.

4. Methodology

The proposed system is an automated system that detects and measures allergic wheals in SPT images by integrating image processing and machine learning techniques. The entire pipeline is deployed as a Streamlit web application and is designed to operate in real time without requiring specialized hardware. The overall workflow of the proposed automated SPT image-analysis system is illustrated in Figure 1.

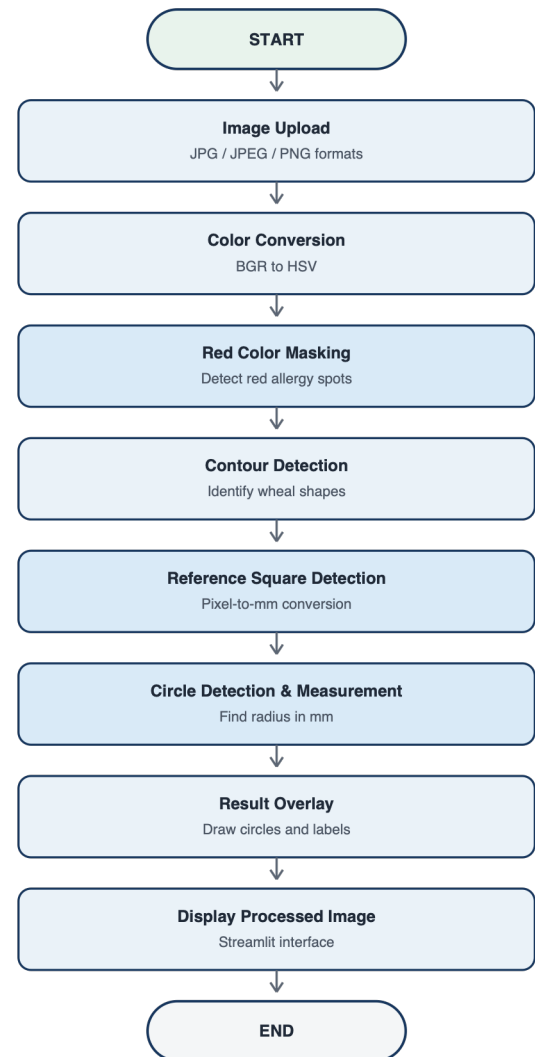


Figure 1. Workflow of the proposed automated SPT image-analysis system.

4.1. Dataset Description

This study used the “Automated Allergy Detection: Skin Prick Test Image” dataset, which is publicly available on Kaggle under the MIT License. A total of 100 Skin Prick Test (SPT) images were used in this study. As all images were obtained from this publicly available source, no participant recruitment, SPT procedures, or new clinical image acquisition were undertaken by the authors.

The publicly available dataset functioned as an additional image resource for the confirmation of the concept. The absence of thorough demographic, clinical, image-acquisition, or annotation metadata in the dataset documentation precluded independent validation of these variables in the present study. As a result, the dataset was perceived as an image-focused evaluation instrument rather than a methodically compiled clinical cohort.

The publicly available dataset was devoid of patient identities, demographic information, and details regarding the number of unique patients included. Therefore, the number and distribution of unique patients could not be independently determined, and patient-level separation of the data could not be verified. The dataset did not provide standardized skin-tone classifications or sufficient demographic information to independently assess the distribution of skin tones. Images contained SPT reaction areas that were used for automatic recognition of wheals and estimation of wheal size.

Since patient identifiers were not provided in the publicly accessible dataset, the number of unique patients and the possibility of repeated images from the same patient could not be independently determined. Consequently, patient-level or test-site-level splitting could not be performed. The dataset was divided at the image level into 70 images (70%) for training and 30 images (30%) for held-out evaluation, with all wheals within a given image assigned to the same subset. The 30-image evaluation subset was excluded from model training and parameter tuning. Because patient-specific identifiers were unavailable, independent verification that photographs from the same patient or testing session were not distributed across the training and evaluation subsets was not possible.

The identity, number, professional background, and clinical specialization of the individuals who created or reviewed the original reference measurements were not disclosed in the publicly accessible dataset or supporting documentation. The annotation process, including whether measurements were independently confirmed by multiple assessors and whether interobserver agreement was evaluated, was also not specified. Therefore, these characteristics could not be independently verified in the present retrospective study. A representative image from the dataset is shown in Figure 2.



Figure 2. Representative SPT image from the publicly available dataset.

4.2. Image Preprocessing

Every uploaded photo was preprocessed to improve the appearance of allergic reactions and lessen the effects of background noise and uneven lighting. The raw image was first transformed into the Hue-Saturation-Value (HSV) color space from the Blue-Green-Red (BGR) color standard that OpenCV uses. HSV separates color and intensity information, making color-based segmentation more robust under different lighting conditions [13]. The binary mask was generated with preset lower and upper HSV thresholds to highlight the red colored areas in the image, which are often correlated with areas with allergic reaction. The masked image was then subjected to Canny edge detection to detect sharp changes in intensity, which could be wheal boundaries [14]. After that, morphological operations including erosion and dilation were performed to remove small artifacts, such as ink markings, shadows, and fine line annotations introduced during clinical testing, while preserving regions which were consistent with wheal morphology [14].

Instead of just picking the largest contour, we combined color and geometric features to define the reference marker. Then the candidate contours are checked for contour area, bounding-box dimensions, aspect ratio, rectangularity, color properties, and expected position in the image after color segmentation. Rectangularity was computed as the ratio of the area of the contour to the area of the bounding box. The marker candidates were selected from contours with features similar to the reference marker and removed from contours with features different from

the reference marker. The reference marker was removed from the image prior to the analysis of the wheal. This was to reduce the chance of misidentification of the reference marker as a large wheal, area of erythema, handwritten mark, or another image artifact.

4.3. Wheal Detection and Measurement

The detected wheal regions were identified using colour-based image segmentation and contour analysis. The preprocessed image was converted from BGR to HSV colour space to separate the reddish wheal regions from the surrounding skin. A colour mask was generated using the selected HSV threshold range, followed by morphological processing to reduce small isolated regions and image noise. Canny edge detection was then applied, and contours were extracted from the processed binary image. Candidate contours were filtered according to their geometric characteristics and minimum contour-area criteria to reduce non-wheal regions and image artefacts. For each retained contour, a minimum enclosing circle was fitted using OpenCV, and the radius was obtained in pixels. The reference marker was used to establish the pixel-to-millimetre conversion factor: $S = L_{mm} / L_{pixel}$, where L_{mm} is the physical length of the reference marker in millimetres, and L_{pixel} is its corresponding measured length in pixels.

The wheal diameter was calculated using the radius of the fitted minimum enclosing circle and the calibration factor: $D_{mm} = 2 \times r_{pixel} \times S$, where r_{pixel} is the fitted-circle radius in pixels, and S is the pixel-to-millimetre conversion factor. The automated measurement therefore measures the diameter of the smallest enclosing circle, rather than the longest wheal diameter or the average of 2 orthogonal diameters. The calculated diameter was shown on the processed image through the Streamlit interface. However, the secondary dataset available in the public domain did not include detailed image-acquisition metadata and reference marker information, and thus the physical nature and consistency of the reference marker could not be independently verified. Furthermore, the information required to quantify or correct for perspective distortion was lacking, and therefore no explicit geometric or homographic correction was performed. Potential measurement errors related to marker uncertainty, perspective, and image-acquisition conditions are acknowledged as limitations.

4.4. Machine Learning Classification

A Support Vector Machine (SVM) classifier was employed to classify the detected wheals according to the predefined

computational severity categories. Geometric features, including radius, area, aspect ratio, and circularity, were extracted from the detected wheal contours and used as input features for the classifier. Based on the wheal-diameter thresholds used in this study, three computational categories were defined: mild (3–4 mm), moderate (>4–8 mm), and severe (> 8 mm). A wheal diameter of > 3 mm was considered a positive SPT reaction according to the available SPT reference used in this study. These thresholds were applied solely as predefined computational categories and should not be interpreted as independently validated clinical severity classifications.

The motivation for including the SVM was to investigate whether combining multiple geometric features, including radius, contour area, aspect ratio, and circularity, could provide a more flexible categorization framework than direct diameter-based thresholding. Direct diameter thresholds remain simpler, more interpretable, and easier to reproduce. Therefore, the present study does not claim that the SVM provides superior classification performance over direct diameter-based categorization; its role is exploratory and intended to support future evaluation using independently established clinical severity labels.

The automated measurements were evaluated against the available reference measurements. The SVM categories were generated using the predefined diameter-based computational thresholds described above; however, independent clinical severity reference labels were not available in the public dataset. However, the publicly available dataset and documentation did not provide sufficient information on the identities, number, credentials, or clinical specializations of the persons who developed or reviewed the initial reference measures. The documentation does not indicate if the assessors were blinded to the automated outcomes, how conflicts were managed, or whether interobserver agreement was established. Consequently, these characteristics could not be independently verified in the present retrospective study. Therefore, the available annotations are referred to as reference measurements, while the SVM outputs are treated as predefined computational categories rather than independently established clinical ground truth.

For reproducibility, the SVM classifier was implemented using a radial basis function (RBF) kernel with $C=1.0$ and $\gamma=\text{scale}$. The geometric features were standardized using StandardScaler before classification. No class weighting was applied, and the default class-weight setting was retained. The SVM hyperparameters were specified a priori

and were not optimized using the held-out evaluation subset.

The SVM analysis used the same image-level training and held-out evaluation split described in Section 4.1.

4.5. Web Application

The full processing pipeline is deployed as an interactive web application built with Streamlit. Users can upload SPT images in JPG, JPEG, or PNG format through a simple drag-and-drop interface. The system processes the uploaded image in real time. The preprocessed image is displayed with annotated wheal contours and measured diameters, together with a summary table of all detected wheal measurements.

The processed image and measurement data can be downloaded directly from the interface for subsequent recording or analysis. The application is designed to run on standard consumer hardware without a graphics processing unit, indicating potential for deployment in outpatient clinics, community health centres, and remote or resource-limited settings, subject to further clinical validation. Best results are achieved when images are captured perpendicular to the skin surface, with consistent lighting and the reference square clearly visible within the frame.

4.6. Experimental Setup

The proposed automated SPT analysis system was implemented in Python and deployed as an interactive web-based application using the Streamlit framework. OpenCV was used for image loading, BGR-to-HSV colour-space conversion, colour-based masking, Canny edge detection, morphological processing, contour extraction, reference-square detection, and geometric measurement [16]. Imutils was used to support image-processing operations, while the Support Vector Machine (SVM) classifier was implemented using the scikit-learn library. The SVM classifier used geometric features extracted from detected contours, including radius, area, aspect ratio, and circularity [15]. NumPy was used for numerical computations and array operations.

The main image-analysis workflow is illustrated in Figure 1. The processing sequence consisted of image upload, BGR-to-HSV colour-space conversion, red-colour masking, contour detection, reference-square-based pixel-to-millimetre calibration, wheal measurement, result annotation, and display of the processed image through the Streamlit interface. The same processing pipeline was applied to all images in the dataset. End-to-end processing

time was measured from image submission to the display of the final processed result.

Final quantitative evaluation was performed exclusively on the 30-image held-out evaluation subset described in Section 4.1; the 70 training images were excluded from the reported quantitative evaluation and were not used for final performance assessment.

5. Results and Discussion

5.1 Overall System Performance

The quantitative evaluation was performed exclusively on the 30-image held-out evaluation subset defined in Section 4.1. The automated measurements showed a mean absolute error (MAE) of 0.48 ± 0.12 mm relative to the available reference measurements. The complete processing pipeline, including image preprocessing, contour detection, reference-marker calibration, wheal measurement, and the exploratory SVM component, required less than two seconds on average per image. The 70 training images were excluded from the reported held-out evaluation results and were not used for final performance assessment. The quantitative results in this work are on the 30-image held-out evaluation subset, rather than on the full 100-image dataset. A representative processed image showing the detected wheal contour and corresponding automated measurement is presented in Figure 3.



Figure 3. Representative processed output showing detected wheal contours and corresponding measurements.

The previously reported detection accuracy of 95.2%, false-positive rate of 3.1%, and false-negative rate of 1.7% were removed because the underlying true-positive (TP), false-positive (FP), false-negative (FN), and true-negative (TN) counts could not be independently verified from the available source dataset. The revised quantitative results are therefore restricted to metrics that can be directly supported by the available evaluation data.

Reference-marker detection formed an essential preprocessing step because the identified marker was used to establish the pixel-to-millimetre conversion required for wheal-diameter measurement. The marker was identified using the colour and geometric criteria described in Section 4.2 rather than solely by contour size. However, a separate quantitative marker-detection success or failure rate could not be independently established because the publicly available dataset did not provide independent marker annotations or sufficient acquisition-level information for systematic validation of the marker-detection step. Therefore, marker-detection performance is acknowledged as a methodological limitation, and the reported measurement results should be interpreted within the calibration constraints of the available images.

The quantitative measurement and processing results obtained from the held-out evaluation subset are summarized in Table 1.

Table 1. Quantitative Performance of the Proposed System

Metric	Mean $\pm$ SD	Description
MAE (mm)	0.48 ± 0.12	Mean absolute error between automated and available reference measurements
Evaluation images (n)	30	Held-out images used for final quantitative evaluation and excluded from training and parameter tuning
Processing Time (s)	< 2	Average end-to-end time per image on standard consumer hardware

For visual comparison, the principal quantitative performance measures are presented in Figure 4. As shown in Figure 4A, the automated wheal-diameter measurement achieved an MAE of 0.48 ± 0.12 mm on the 30-image held-out evaluation subset relative to the available reference measurements. The relatively small variation in measurement error indicates that the automated measurements remained reasonably consistent across the evaluated images. Figure 4B summarizes the computational efficiency of the system, showing that the complete end-to-end processing pipeline required less than 2.0 s per image. This processing time includes image preprocessing, contour detection, reference-marker calibration, wheal measurement, and the exploratory SVM component. Together, these results demonstrate that the proposed system can provide automated wheal measurements with low measurement error while maintaining relatively fast image processing.

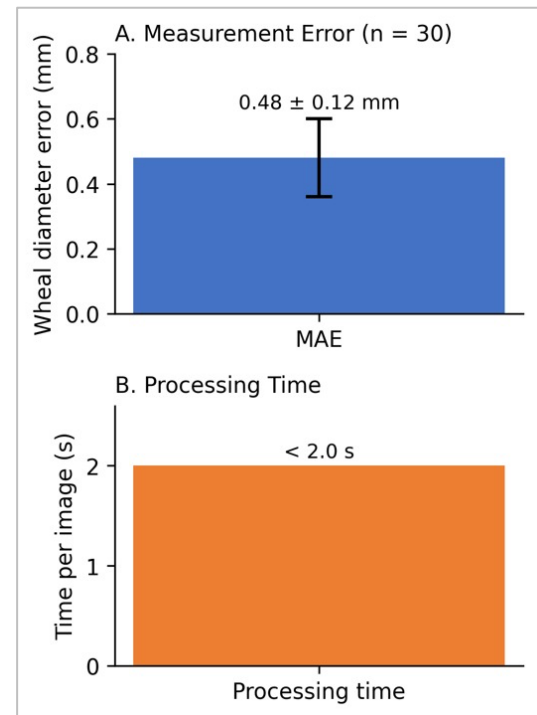


Figure 4. Measurement error and processing time of the proposed system.

5.2. Clinical Significance of the Error Margin

The mean absolute error of 0.48 mm should be interpreted in the context of clinical SPT interpretation. A positive SPT reaction is commonly defined using a wheal-size criterion relative to the negative control [3], [5]. The observed MAE indicates that, on average, the automated measurements were close to the reference measurements used for evaluation. Manual SPT measurement is subject to variability arising from differences in wheal delineation and measurement methods [3], [5]. The low measurement error observed in the present study therefore supports the potential of the automated system to provide objective and consistent wheal measurements.

The low standard deviation of ± 0.12 mm further indicates limited variability in the measurement errors observed across the evaluated images. The results show that the system has produced consistent measurements for the present dataset. However, before drawing the conclusion that the system is more standardised or reproducible than manual SPT measurement in routine clinical practice, further validation in larger and more diverse clinical datasets, imaging conditions, skin tones, and reaction sizes is needed.

The root mean square error (RMSE) was not reported because the individual paired reference and automated measurements were not retained in the available source dataset. Aggregate MAE and standard deviation values are

insufficient to reconstruct the individual paired observations required for a valid RMSE calculation. Reporting a reconstructed RMSE would therefore introduce unsupported assumptions.

A reference-versus-automated measurement plot was also not generated because the individual paired reference and automated measurement values were not retained in the available source dataset. Such a plot requires paired measurements for each evaluated wheal or image and cannot be validly reconstructed from aggregate MAE and standard deviation values. Future validation should retain paired reference and automated measurements for every evaluated wheal and report a reference-versus-automated scatter plot together with appropriate agreement and error analyses.

5.3. Detection Accuracy and Error Analysis

The evaluation of detection performance requires independent ground-truth annotations for individual wheals and non-wheal regions to determine true-positive (TP), false-positive (FP), false-negative (FN), and true-negative (TN) counts. The underlying annotations required to independently reconstruct these counts were not available in the source dataset used in this study. Therefore, the previously reported detection accuracy of 95.2%, false-positive rate (FPR) of 3.1%, and false-negative rate (FNR) of 1.7% have been removed from the revised manuscript. No TP, FP, FN, or TN values are reported because these values cannot be reliably reconstructed from the available data without introducing unsupported assumptions. For a fully annotated future evaluation, the metrics would be calculated as follows:

$$\begin{aligned}\text{Accuracy} &= (TP + TN) / (TP + TN + FP + FN) \\ \text{FPR} &= FP / (FP + TN) \\ \text{FNR} &= FN / (FN + TP)\end{aligned}$$

The available images nevertheless indicate several potential sources of detection error associated with the image-processing procedure. Faint pencil or pen markings may partially overlap with the red colour range used for HSV-based masking, while small areas of skin erythema near the test site may produce regions with visual characteristics similar to those of wheals. Such regions may therefore be retained as candidate contours during segmentation. Conversely, very small wheals, particularly those with diameters below approximately 2 mm, or wheals with a more pink than red appearance, may be less responsive to the selected HSV colour range and may subsequently be excluded by the minimum contour-area filtering step.

These observations are qualitative and are not based on an independently annotated error-category dataset. They therefore should not be interpreted as quantitative false-positive or false-negative rates. They instead identify potential sources of error that can be investigated in future validation studies. Possible improvements include adaptive adjustment of the lower HSV colour threshold and evaluation of a smaller or context-dependent minimum contour-area threshold for borderline wheal reactions.

The adaptive thresholding approach showed improved qualitative behaviour compared with the static thresholding approach in images affected by uneven or low illumination. However, the publicly available dataset does not provide standardized image-acquisition metadata or predefined lighting categories. Consequently, a controlled quantitative comparison of detection performance under different illumination conditions could not be performed. Future studies should use independently annotated images acquired under standardized and documented imaging conditions to quantify the effect of adaptive thresholding on detection performance.

5.4. Exploratory Diameter-Based Categorization

The SVM classifier was implemented as an exploratory computational component to categorize detected wheals according to predefined diameter-based computational categories. The categories were defined as mild (3–4 mm), moderate (>4–8 mm), and severe (>8 mm) based on the diameter thresholds used in this study.

The publicly available dataset did not provide independently established mild, moderate, or severe reference labels. Therefore, independent clinical severity-classification accuracy could not be calculated from the available dataset. The resulting categories should consequently be interpreted as predefined computational groupings rather than clinically validated severity classifications.

The SVM analysis therefore serves as an exploratory component of the proposed image-analysis workflow rather than evidence of clinically validated severity classification. Future evaluation should use an independently annotated dataset containing clinically established severity categories and should report a confusion matrix, precision, recall, specificity, F1-score, and macro- and weighted-average performance measures.

5.5. Comparison with Prior Work

The proposed system extends previous work on biomedical image segmentation and objective SPT

measurement. Ronneberger et al. [17] introduced the U-Net architecture for precise biomedical image segmentation, while Konstantinou et al. [18] identified the longest wheal diameter as an appropriate parameter for evaluating SPT reactions. Previous computer-assisted methods relied on scanning programs for the estimation of SPT reactions [19]. Poulsen et al. [20] validated a scanning device and computer software for a correct measurement of skin-prick-test area. Conversely, the present system performs automated wheal analysis directly from digital images, achieves an MAE of 0.48 mm, and incorporates exploratory diameter-based categorization through a web-based interface.

Characteristics of the image and procedural factors can also influence the assessment of SPT reactions. Holm et al. [21] investigated a colour-enhancement method to improve the recognizability of histamine skin reactions, and Andersen et al. [22] found that the weight of the lancet can affect the resulting wheal diameter. Conversely, the proposed system integrates digital image preprocessing, automated wheal measurement, and an exploratory diameter-based SVM component within a single workflow. It employs classical image processing rather than deep learning, which has practical advantages such as interpretability, reduced computational requirements, the ability to run on low-cost hardware, and the ability to deploy without GPU infrastructure.

5.6. Limitations

Several limitations of the current study should be acknowledged. First, the proposed system was developed and evaluated using images obtained from a publicly available dataset. Although the dataset was suitable for proof-of-concept evaluation, the findings should be validated using additional independent datasets before broader clinical conclusions are drawn. Future studies should evaluate the system across more diverse skin tones, imaging devices, lighting conditions, allergen panels, and clinical environments.

Another issue was the lack of confirmation of patient-level independence between training and evaluation subsets. The dataset was divided at the image level, and all wheals in an image were preserved in the same subset; however, patient identifiers were not accessible to verify that photographs from the same patient or test session were not included in both subsets. Thus, patient-level data leakage cannot be precluded entirely, and the reported evaluation performance should be considered as an image-level evaluation and not as an independently confirmed patient-level evaluation.

Second, the authors did not control the original image-acquisition process because the images were obtained from a secondary public source. Variations in image quality, illumination, camera position, and other acquisition conditions may therefore influence system performance. Prospective multicenter validation using standardized and independently collected clinical data will be required to assess generalizability in real-world clinical settings.

5.7 Clinical and Deployment Implications

The system is designed to support, rather than replace, clinical assessment by trained healthcare professionals. It provides automated quantitative measurements that could potentially facilitate standardized recording and comparison of SPT wheal sizes.

The web-based implementation and relatively low computational requirements indicate potential for use in resource-limited environments. However, the present study represents a proof-of-concept evaluation based on a publicly available image dataset. Clinical deployment, integration with electronic health records, and use during routine consultations would require prospective validation using independently collected clinical data and standardized image-acquisition procedures. Therefore, the current findings demonstrate technical feasibility rather than established clinical suitability.

6. Conclusion

An automated image-analysis system for the detection and quantification of wheal responses in skin prick test (SPT) images was developed and evaluated. The proposed workflow combines HSV colour segmentation, Canny edge detection, morphological image processing, reference-marker-based pixel-to-millimetre calibration, contour analysis, and an exploratory Support Vector Machine (SVM) component within a Streamlit web application. On the 30-image held-out evaluation subset, the system achieved a mean absolute error of 0.48 ± 0.12 mm relative to the available reference measurements, with an average end-to-end processing time of less than two seconds per image. These findings demonstrate the technical feasibility of automated wheal measurement using standard digital SPT images. However, the results should be interpreted as a proof-of-concept evaluation rather than evidence of clinical validity or diagnostic accuracy. The available dataset did not provide sufficient information to independently verify patient-level independence, the original annotation process, clinical assessor characteristics, independent wheal-level detection annotations, or clinically established severity categories.

Consequently, detection metrics requiring independently verified TP, FP, FN, and TN counts and clinically validated SVM classification performance could not be established from the available data. Therefore, the exploratory SVM component should not be interpreted as a clinically validated severity-classification system. Future studies should evaluate the system using larger, independently collected and clinically annotated datasets with patient-level separation, standardized image acquisition, independently established severity labels, and paired reference and automated measurements to enable comprehensive detection, measurement-error, and classification analyses.

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable.

Author Contributions: Z.P. contributed to conceptualization, methodology, software, validation, formal analysis, investigation, visualization, and writing—original draft. S.D. contributed to conceptualization, methodology, supervision, validation, and writing—review and editing. Both authors read and approved the final manuscript.

Data Availability Statement: The dataset used in this study was obtained from a publicly accessible Kaggle repository and is available at: <https://www.kaggle.com/datasets/xiaoyazai/skin-allergy-detection>.

Code Availability: The source code developed for the automated Skin Prick Test (SPT) image-analysis system is publicly available at <https://github.com/Xiao-Tu-SD/Skin-Allergy-Detection.git>.

Conflicts of Interest: The authors declare no conflicts of interest.

AI Use Declaration: During the preparation of this manuscript, the authors used Grammarly and ChatGPT solely to improve English language, grammar, readability, and clarity. These tools were not used to generate research data, perform analyses, interpret results, or formulate scientific conclusions. The authors reviewed and edited all AI-assisted language modifications and take full responsibility for the content of the manuscript.

References

1. Shin YH, Hwang J, Kwon R, Lee SW, Kim MS, et al. Global, regional, and national burden of allergic disorders and their risk factors in 204 countries and territories, from 1990 to 2019: A systematic analysis for

- the Global Burden of Disease Study 2019. *Allergy*. 2023;78(8):2232–2254. doi:10.1111/all.15807.
2. Sicherer, S.H.; Sampson, H.A. Food allergy: Epidemiology, pathogenesis, diagnosis, and treatment. *J. Allergy Clin. Immunol.* 2014, 133, 291–307.e5. <https://doi.org/10.1016/j.jaci.2013.11.020>.
3. Heinzerling, L.; Mari, A.; Bergmann, K.C.; Bresciani, M.; Burbach, G.; Darsow, U.; Durham, S.; Fokkens, W.; Gjomarkaj, M.; Haahtela, T.; Todo Bom, A.; Wöhr, S.; Maibach, H.; Lockey, R. The skin prick test: European standards. *Clin. Transl. Allergy* 2013, 3, 3. <https://doi.org/10.1186/2045-7022-3-3>.
4. Bernstein, I.L.; Li, J.T.; Bernstein, D.I.; Hamilton, R.; Spector, S.L.; Tan, R.; Sicherer, S.; Golden, D.B.K.; Khan, D.A.; Nicklas, R.A.; Portnoy, J.M.; Blessing-Moore, J.; Cox, L.; Lang, D.M.; Oppenheimer, J.; Randolph, C.C.; Schuller, D.E.; Tilles, S.A.; Wallace, D.V.; Levetin, E.; Weber, R. Allergy diagnostic testing: An updated practice parameter. *Annals of Allergy, Asthma & Immunology*, 2008, 100, S1–S148. [https://doi.org/10.1016/S1081-1206\(10\)60305-5](https://doi.org/10.1016/S1081-1206(10)60305-5).
5. van der Valk, J.P.M.; Gerth van Wijk, R.; Hoorn, E.; Groenendijk, L.; Groenendijk, I.M.; de Jong, N.W. Measurement and interpretation of skin prick test results. *Clin. Transl. Allergy* 2016, 6, 8. <https://doi.org/10.1186/s13601-016-0092-0>.
6. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.W.M.; van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* 2017, 42, 60–88. <https://doi.org/10.1016/j.media.2017.07.005>.
7. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017, 542, 115–118. <https://doi.org/10.1038/nature21056>.
8. Wöhr, S.; Vigl, K.; Binder, M.; Stingl, G.; Prinz, M. Automated measurement of skin prick tests: An advance towards exact calculation of wheal size. *Exp. Dermatol.* 2006, 15, 119–124. <https://doi.org/10.1111/j.1600-0625.2006.00388.x>.
9. Pineda, J.; Vargas, R.; Romero, L.A.; Marrugo, J.; Meneses, J.; Marrugo, A.G. Robust automated reading of the skin prick test via 3D imaging and parametric surface fitting. *PLoS ONE* 2019, 14, e0223623. <https://doi.org/10.1371/journal.pone.0223623>.
10. Justo, X.; Diaz, I.; Gil, J.J.; Gastaminza, G. Medical device for automated prick test reading. *IEEE J. Biomed. Health Inform.* 2018, 22, 895–903. <https://doi.org/10.1109/JBHI.2017.2680840>.
11. Lee, Y.H.; Shim, J.S.; Kim, Y.J.; Jeon, J.S.; Kang, S.Y.; Lee, S.P.; Lee, S.M.; Kim, K.G. Allergy wheal and erythema segmentation using Attention U-Net. *J. Imaging Inform. Med.* 2025, 38, 467–475. <https://doi.org/10.1007/s10278-024-01075-0>.
12. Ansotegui, I.J.; Melioli, G.; Canonica, G.W.; Caraballo, L.; Villa, E.; Ebisawa, M.; et al. IgE allergy diagnostics and other relevant tests in allergy: A World Allergy Organization position paper. *World Allergy Organ. J.* 2020, 13, 100080. <https://doi.org/10.1016/j.waojou.2019.100080>.
13. Chavolla, E.; Zaldivar, D.; Cuevas, F.; Rayas-Sánchez, J.E. Improved Unsupervised Color Segmentation Using a Modified HSV Color Model and a Bagging Procedure in K-Means++ Algorithm. *Mathematical Problems in Engineering* 2018, 2018, 2786952. <https://doi.org/10.1155/2018/2786952>.
14. Canny, J. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 1986, 8, 679–698. <https://doi.org/10.1109/TPAMI.1986.4767851>.
15. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* 1995, 20, 273–297. <https://doi.org/10.1007/BF00994018>.

16. Maglogiannis, I.G.; Zafiroopoulos, E.P. Characterization of digital medical images utilizing support vector machines. *BMC Med. Inform. Decis. Mak.* **2004**, *4*, 4. <https://doi.org/10.1186/1472-6947-4-4>.
17. Ronneberger, O.; Fischer, P.; Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28
18. Konstantinou, G.N.; Bousquet, P.J.; Zuberbier, T.; Papadopoulos, N.G. The longest wheal diameter is the optimal measurement for the evaluation of skin prick tests. *Int. Arch. Allergy Immunol.* **2010**, *151*, 343–345. <https://doi.org/10.1159/000250443>.
19. Pijnenborg, H.; Nilsson, L.; Dreborg, S. Estimation of skin prick test reactions with a scanning program. *Allergy* **1996**, *51*, 782–788. <https://doi.org/10.1111/j.1398-9995.1996.tb04467.x>.
20. Poulsen, L.K.; Liisberg, C.; Bindslev-Jensen, C.; Malling, H.J. Precise area determination of skin-prick tests: Validation of a scanning device and software for a personal computer. *Clin. Exp. Allergy* **1993**, *23*, 61–68. <https://doi.org/10.1111/j.1365-2222.1993.tb02485.x>.
21. Holm, O.; Lindell, E.; Malm, L. Colour change in the human histamine wheal made visible by LYNN: A technique to enhance colour differences. *Skin Res. Technol.* **2010**, *16*, 385–389. <https://doi.org/10.1111/j.1600-0846.2010.00457.x>.
22. Andersen, H.H.; Lundgaard, A.C.; Petersen, A.S.; Hauberg, L.E.; Sharma, N.; Hansen, S.D.; Elberling, J.; Arendt-Nielsen, L. The lancet weight determines wheal diameter in response to skin prick testing with histamine. *PLoS ONE* **2016**, *11*, e0156211. <https://doi.org/10.1371/journal.pone.0156211>.

Disclaimer: All views, interpretations, and data presented in Impaxon publications are the sole responsibility of the respective authors. These do not necessarily reflect the opinions of Impaxon or its editorial team. Impaxon and its editors assume no liability for any harm or loss arising from the use of information, procedures, or materials discussed in the published content.

Publisher’s Note: Impaxon remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.