E-ISSN: 2963-4946
Vol. 1 No. 6 March 2023, pp. 288-306
DOI: 10.59888/ajosh.v4i10.736

Asian Journal of Social and Humanities

Evaluation of Saliency-Guided Foreground Extraction for Medicinal Plant Flower Classification on the PhulMedic Dataset

Viandre Chika Baratama
Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) IKMI Cirebon, Indonesia
Corresponding Author: viandrechikabaratama@gmail.com

Keywords:
phulmedic; flower classification; saliency detection; foreground extraction; mobileNetV3Small

This is an open access article under the CC BY-SA license.

Copyright © 2026 by Author. Published by Pelopor Publikaasi Akademika

Abstract

Background: Medicinal plant flower image classification faces variations in background, lighting, and visual characteristics that can affect the feature representations of deep learning models, yet the benefit of background removal remains unclear.
Objective: This study evaluates whether Saliency-Guided Foreground Extraction improves MobileNetV3Small classification on the PhulMedic dataset and what computational cost it adds.
Method: The dataset contains 5,140 images of Crape Jasmine, Hibiscus, Madagascar Periwinkle, Marigold, and Wedelia, split with stratification at a 70:15:15 ratio. Original images were compared with foreground images produced by Static Saliency Spectral Residual, a threshold of 25, and morphological closing with a 7 × 7 elliptical kernel. An ImageNet-pretrained MobileNetV3Small with a frozen backbone and 224 × 224 inputs was trained under identical settings, and compared using the exact McNemar test.
Results: On 771 test images, the original condition achieved 100% accuracy and macro-F1, whereas the foreground condition achieved 99.61% for both, with three misclassifications; the difference was not statistically significant ($p = 0.25$). Foreground preprocessing plus training was 20.76% slower than baseline training.
Conclusion: Foreground extraction produced no detectable difference in performance while adding computational overhead. Because the baseline was saturated and only one split was used, this design cannot show whether foreground extraction helps.

INTRODUCTION

Plant identification from digital images is an application of computer vision that can support biodiversity documentation and the preliminary recognition of species (Chen et al., 2023). Flowers have visual characteristics such as shape, colour, petal arrangement, and texture that can be used to distinguish categories. In the context of medicinal plants, image classification can serve as a tool to support species recognition (Malik et al., 2022), but it does not replace formal botanical identification or the testing of medicinal properties and chemical content.

Medicinal plants continue to play an important role in primary health care, traditional medicine, and the supply of raw materials for herbal and pharmaceutical products in many countries, including Bangladesh and Indonesia, where a large number of species are used by local communities. Correct identification is a prerequisite for safe use, because morphologically similar flowers may belong to species with different medicinal

properties, and some flowers are poisonous while only a few have medicinal value (Huang & Xu, 2023). Manual identification depends on botanical expertise that is not always available in the field, which motivates the development of automatic, low-cost recognition tools that can run on mobile devices (Malik et al., 2022).

Convolutional neural networks (CNNs) learn visual features hierarchically and have been widely used for plant image classification (Alzubaidi et al., 2021). Transfer learning allows pretrained models to be used when data and computational resources are limited. Flower classification studies have used a range of CNN-based approaches, including species localization followed by recognition (Abbas et al., 2022), comparisons of CNN architectures (Rajkomar & Pudaruth, 2023), lightweight networks with multi-scale feature fusion and attention (Zeng et al., 2023), selective-attention hybrids (Cao & Song, 2021), and lightweight transfer-learning models for flower images (Fei et al., 2023).

Medicinal flower and plant classification has also been addressed with CNNs (Huang & Xu, 2023; Roopashree & Anitha, 2021), and dedicated image datasets have been proposed to support medicinal flower classification (Huang et al., 2021). Reported results in medicinal flower classification are generally high; for example, Huang and Xu (2023) obtained an accuracy of 97.64% and an F1-score of 97.82% on Chinese medicinal flowers with a ResNet101 model tuned by Bayesian optimization. Such results, however, are usually obtained from a single train–test partition, and near-ceiling scores make it difficult to determine whether an additional preprocessing step contributes any measurable benefit.

MobileNetV3Small was chosen for this study because it is a lightweight CNN architecture obtained through hardware-aware architecture search (Howard et al., 2019) and is suitable for the CPU-constrained experiments conducted in Google Colab CPU. The small variant of the same architecture has also been applied to plant image classification on edge computing devices (Khan et al., 2023). Its small computational footprint also makes it a realistic choice for applications on mobile devices, where preprocessing costs become as important as classification accuracy.

Flower classification in natural environments is challenged by background variation. Leaves, soil, vegetation, hands, and other objects can appear together with flowers. The background can provide helpful context, but it can also become an irrelevant shortcut if the correlation between background and label is not consistent. Moayeri et al. (2022) showed that model sensitivity to foregrounds, backgrounds, and visual attributes needs to be considered when evaluating classifiers. Studies on background removal in plant images also show that the effect of removing the background depends on the data context and on the quality of the removal process (KC et al., 2021).

Segmentation methods designed for cluttered backgrounds, such as unsupervised leaf segmentation Guo et al., (2023) and local refinement mechanisms (Ma et al., 2023), illustrate how difficult it is to separate plant parts reliably from their surroundings. Saliency detection attempts to highlight visual regions that are considered interesting or conspicuous. The spectral residual approach detects saliency from the frequency domain without training (Hou & Zhang, 2007), whereas modern salient object detection

increasingly relies on deep networks (Borji et al., 2019; Wang et al., 2022), including models that jointly learn salient-object, foreground-contour, and edge cues (Wu et al., 2019). Comparative evaluations show that the performance of salient object detection techniques varies considerably across images and conditions (Ahmed et al., 2022).

Moreover, saliency is not semantic segmentation: salient regions are not necessarily flowers, and parts of a flower that are not conspicuous may be ignored. Foreground extraction can therefore reduce irrelevant background, but it can also remove discriminative information. The PhulMedic dataset provides five classes of medicinal plant flowers (PhulMedic: A Bangladeshi Medicinal Flower Dataset (Version 2), 2025). Exploratory data analysis (EDA) showed variation in brightness, contrast, RGB colour, blur, and background. This study therefore tests empirically whether saliency-based foreground extraction improves MobileNetV3Small classification and assesses its effect on accuracy and computational cost. The experiment is designed as a controlled comparison in which only the image preprocessing differs between the two conditions.

To the best of the author's knowledge, published work on PhulMedic has so far focused mainly on the release and description of the dataset (PhulMedic: A Bangladeshi Medicinal Flower Dataset (Version 2), 2025), and no study has reported the effect of saliency-based background removal on this dataset. Earlier work on plant images has shown that the effect of background removal depends on the data context (KC et al., 2021) and that classifiers may rely on both foreground and background cues (Moayeri et al., 2022). The novelty of this study therefore lies in a controlled, cost-aware evaluation of training-free spectral residual foreground extraction combined with a lightweight MobileNetV3Small on PhulMedic, including a paired statistical test of the difference between conditions and an analysis of runtime overhead.

The hypothesis of this study is that, because PhulMedic images show heterogeneous backgrounds and lighting, suppressing non-salient background regions reduces the model's reliance on contextual shortcuts and therefore improves, or at least maintains, classification performance. The competing expectation is that coarse, training-free saliency masks remove discriminative flower parts or useful context and thus reduce performance. The expected benefit of this study is evidence-based guidance on whether a low-cost preprocessing step is worth its computational overhead in lightweight medicinal flower recognition systems intended for mobile or resource-constrained deployment, together with a methodological illustration of how a saturated baseline limits the interpretation of preprocessing experiments.

METHOD

Dataset and Exploratory Data Analysis

Version 2 of the PhulMedic dataset (PhulMedic: A Bangladeshi Medicinal Flower Dataset (Version 2), 2025) contains 5,140 images in five classes (Table 1). All images recorded in the notebook have a resolution of 1024 × 1024 pixels with three colour channels. The class distribution is relatively balanced, with a ratio of 1.08:1 between the largest and smallest classes. Wedelia has the most images (1,080) and Hibiscus the fewest

(1,000), a difference of 80 images or about 1.56% of the dataset. With this distribution, macro-precision, macro-recall, and macro-F1 remain relevant for assessing performance in each class. The EDA reported in this subsection is descriptive: it documents the visual heterogeneity (brightness, contrast, colour, and background) that motivates the hypothesis of this study and provides the context in which the effect of foreground extraction and the misclassifications are interpreted.

Table 1. Class distribution of the PhulMedic dataset

Class	Number of images	Percentage
Crape Jasmine	1,020	19.84%
Hibiscus	1,000	19.46%
Madagascar Periwinkle	1,020	19.84%
Marigold	1,020	19.84%
Wedelia	1,080	21.01%
Total	5,140	100%

Source: Data Processed

The EDA recorded the lowest mean brightness for Crape Jasmine (82.05), the highest brightness standard deviation for Wedelia (25.66), and the lowest mean contrast for Wedelia (36.97). The closest pair of classes in mean RGB was Madagascar Periwinkle–Wedelia, with a distance of 15.76. RGB statistics were computed over the whole frame and cannot be regarded as the colour of the flowers alone. These class-level statistics are summarized in Figure 1.

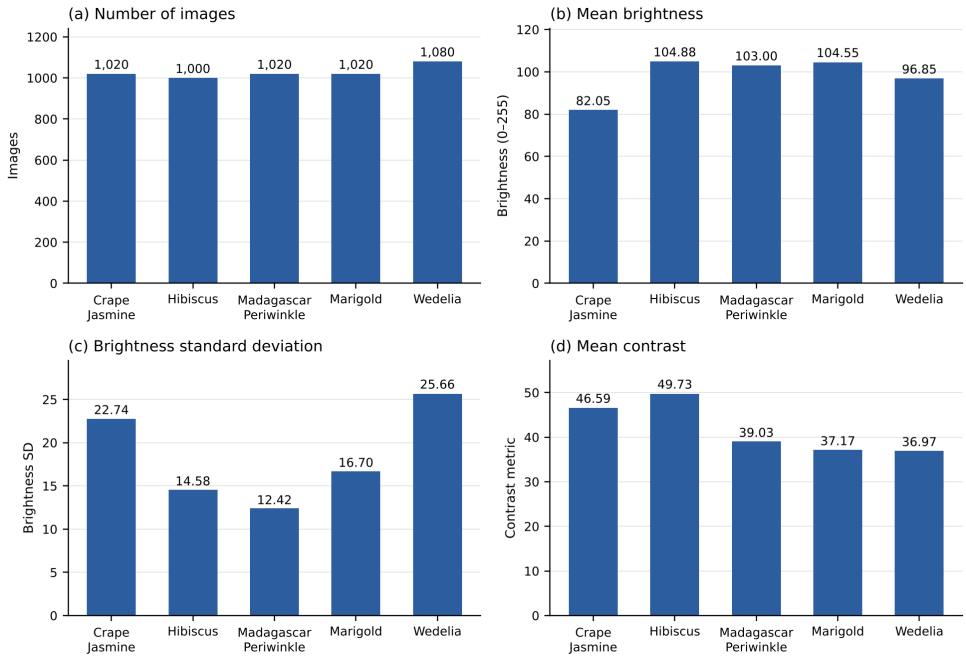


Figure 1. Class-level EDA summary: (a) number of images, (b) mean brightness, (c) brightness standard deviation, and (d) mean contrast

Resolution and Aspect Ratio

All classes have a single resolution of 1024 × 1024 pixels and an aspect ratio of 1.0, with no variation (Table 2). Because the dataset is square and uniform, resizing to 224 × 224 does not require padding to preserve the aspect ratio. Resizing nevertheless reduces spatial detail, especially small details of the petals and pollen.

Table 2. Resolution and Aspect Ratio

Parameter	EDA value
Mean resolution per class	1024 × 1024
Unique resolutions per class	1
Aspect ratio per class	1.0
Aspect ratio min–max	1.0–1.0
Aspect ratio standard deviation	0

Source: Data Processed

Brightness

Crape Jasmine had the lowest mean brightness (82.05) with a standard deviation of 22.74 (Table 3). Wedelia had the highest brightness standard deviation (25.66) and a maximum of 168.30, indicating the most varied range of intra-class lighting. Hibiscus had the highest mean brightness (104.88). No class maximum exceeded 200, but the Crape Jasmine minimum of 21.64 indicates the presence of very dark samples. The 50/200 thresholds mentioned in the EDA are rules of thumb, not a ground truth of image quality.

Table 3. Brightness statistics per class

Class	Mean	SD	Min	Max
Crape Jasmine	82.05	22.74	21.64	139.64
Hibiscus	104.88	14.58	67.19	146.92
Madagascar Periwinkle	103.00	12.42	63.74	138.65
Marigold	104.55	16.70	62.51	146.79
Wedelia	96.85	25.66	36.11	168.30

Source: Data Processed

Contrast

Hibiscus had the highest mean contrast (49.73), whereas Wedelia had the lowest (36.97), followed by Marigold (37.17) (Table 4). Madagascar Periwinkle had a contrast standard deviation of 10.45 and Wedelia 10.87, indicating fairly high within-class variation in contrast. Values below 40 were flagged in the EDA as an indication of low contrast, so Madagascar Periwinkle, Marigold, and Wedelia were at or below this range on average. The contrast metric in the EDA is not a universal standard, so its threshold must be interpreted with caution.

Table 4. Contrast statistics per class

Class	Mean contrast	SD	Min	Max
Crape Jasmine	46.59	8.97	26.21	66.85
Hibiscus	49.73	8.53	26.95	72.24
Madagascar Periwinkle	39.03	10.45	11.98	62.20
Marigold	37.17	5.56	25.14	63.77
Wedelia	36.97	10.87	13.80	57.89

Source: Data Processed

RGB Analysis

The green channel was higher than the red and blue channels for Hibiscus, Madagascar Periwinkle, Marigold, and Wedelia (Table 5 and Figure 2). This may reflect a mixture of green foliage and background rather than petal colour alone. For Hibiscus, G = 117.49, R = 103.40, and B = 43.92; for Marigold, G = 115.87, R = 98.93, and B = 61.15. Crape Jasmine was relatively balanced across RGB (R = 80.87; G = 83.14; B = 79.53), but these means are far below the intensity of pure white because the statistics cover the whole image, including the background. Global colour can therefore be a cue, but it can also be an unstable shortcut. Colour augmentation such as hue and saturation jitter could be considered, with moderate intensity so that the biological colour of the flowers is not distorted.

Table 5. Mean RGB channel intensity per class

Class	R mean	G mean	B mean
Crape Jasmine	80.87	83.14	79.53
Hibiscus	103.40	117.49	43.92
Madagascar Periwinkle	107.75	105.39	78.21
Marigold	98.93	115.87	61.15
Wedelia	94.21	103.33	70.41

Source: Data Processed

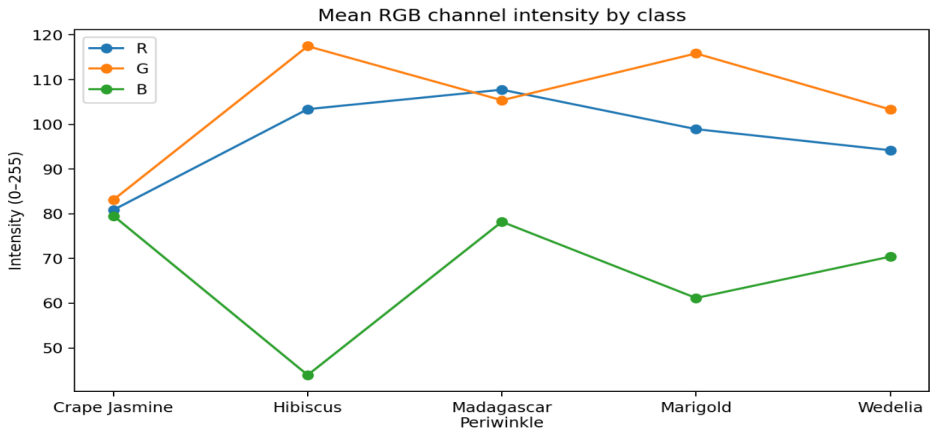


Figure 2. Mean RGB channel intensity per class

Pixel Distribution and Brightness Outliers

Brightness outliers were identified with the interquartile range (IQR) rule, in which values below Q1 − 1.5 × IQR or above Q3 + 1.5 × IQR are flagged (Table 6). The highest number of brightness outliers was in Crape Jasmine (55; 5.39%), followed by Wedelia (24; 2.22%). IQR-based outliers are not automatically corrupted data; they may represent extreme lighting or different scenes. Valid samples should be retained so that the model learns real-world variation, while augmentation and robustness testing can be considered.

Table 6. Brightness distribution and IQR-based outliers per class

Class	Q1	Q3	IQR	Lower bound	Upper bound	Outliers (n)	Outliers (%)
Crape Jasmine	65.27	91.26	25.99	26.29	130.24	55	5.39
Hibiscus	95.55	115.67	20.12	65.37	145.84	3	0.30
Madagascar Periwinkle	94.74	111.37	16.63	69.80	136.32	5	0.49
Marigold	91.73	117.61	25.88	52.91	156.44	0	0.00
Wedelia	77.74	110.88	33.14	28.04	160.59	24	2.22

Source: Data Processed

Duplicates

The notebook identified two duplicate pairs in Crape Jasmine, namely CJ_0241–CJ_0242 and CJ_0249–CJ_0250 (Table 7). It must be checked whether these pairs were placed in different subsets, because data leakage can affect the estimate of test performance. Because the two pairs have consecutive file names, they may originate from burst or repeated shots of the same flower, so exact duplicates may represent only part of the near-duplicate content. A perceptual-hash near-duplicate audit and a group-aware split were not performed in the present experiment and are therefore treated as a limitation.

Table 7. Duplicates detected per class

Class	Duplicates detected	Unique images	Percentage
Crape Jasmine	2	1,018	0.20%
Hibiscus	0	1,000	0%
Madagascar Periwinkle	0	1,020	0%
Marigold	0	1,020	0%
Wedelia	0	1,080	0%

Source: Data Processed

RGB Similarity between Classes

The closest pair in mean RGB was Madagascar Periwinkle–Wedelia (15.76), followed by Marigold–Wedelia (16.28) and Hibiscus–Marigold (17.87) (Table 8). For two classes with mean colour vectors μ_a and μ_b, the distance was computed as in Equation (1). These results are consistent with potential colour confusion, but the global mean RGB

distance is not a measure of similarity in shape or texture, nor of full classification difficulty.

$$d(a, b) = \sqrt{[(R_a - R_b)^2 + (G_a - G_b)^2 + (B_a - B_b)^2]} \quad (1)$$

Table 8. Euclidean distance between class mean RGB vectors

Class pair	Euclidean RGB distance
Madagascar Periwinkle – Wedelia	15.76
Marigold – Wedelia	16.28
Hibiscus – Marigold	17.87
Madagascar Periwinkle – Marigold	21.88
Crape Jasmine – Wedelia	25.86
Hibiscus – Wedelia	31.41
Crape Jasmine – Madagascar Periwinkle	34.92
Hibiscus – Madagascar Periwinkle	36.62
Crape Jasmine – Marigold	41.65
Crape Jasmine – Hibiscus	54.36

Source: Data Processed

Spatial Distribution

The EDA indicated that objects tend to be concentrated in the centre of the frame (Figure 3). This can create a position bias (centre bias). Random translation or shifting, light cropping, and scale augmentation could help, but care must be taken that flowers are not excessively cropped. The Hibiscus map is more diffuse than those of the other classes, suggesting that flowers in this class appear in more varied positions or sizes.

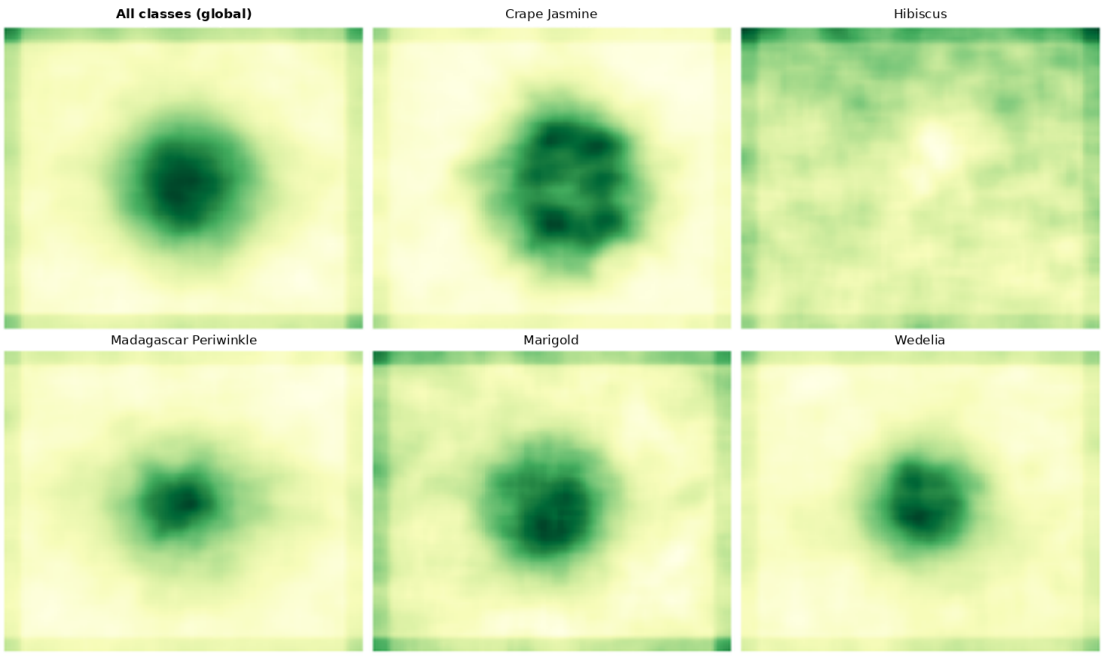


Figure 3. Dominant spatial distribution of objects for all classes and for each class, from the EDA. Darker green indicates locations where flower objects appear more often

Resizing and Model Preprocessing

The source images are 1024 × 1024 × 3, whereas the experimental model uses 224 × 224 × 3 inputs. The number of pixels per image therefore decreases from 1,048,576 to 50,176, a reduction of about 95.21%. Because the source aspect ratio is 1:1, direct resizing does not change the proportions. Resizing reduces the computational load, but small details may be lost. Sample images from the EDA are shown in Figure 4.



Figure 4. Sample flower images from the EDA output with their overexposure scores (not resized from the original dataset files)

Dataset Splitting

The data were split with stratification using random_state = 42 (Table 9). The same split was used in both experiments. Because the dataset contains detected duplicate pairs, the results should be regarded as provisional until duplicate leakage has been audited. Because a single stratified split was used and the baseline reached 100% test accuracy, the evaluation is subject to a ceiling effect: the protocol can reveal a decrease caused by foreground extraction but not an improvement. For this reason, the comparison was complemented by exact confidence intervals for test accuracy and a paired exact McNemar test (see Statistical Analysis).

Table 9. Dataset split

Subset	Number of images	Proportion
Training	3,598	70%
Validation	771	15%
Test	771	15%

Source: Data Processed

Saliency-Guided Foreground Extraction

Foreground extraction used StaticSaliencySpectralResidual from OpenCV, which implements the spectral residual approach of Hou and Zhang (2007). Conceptually, the method computes saliency from the residual of the log-amplitude spectrum with respect to its local average. The image is mapped to the frequency domain, its amplitude and phase are computed, the spectral residual is obtained, and the residual is returned to the spatial domain through the inverse Fourier transform.

For a grayscale image I(x, y), the two-dimensional Fourier transform is written as in Equation (2), where A(u, v) = |F(u, v)| is the amplitude spectrum and φ(u, v) is the phase. The OpenCV implementation computes the saliency map on a downscaled copy of the input image; with the default settings used in this study, the working resolution is 64 × 64

pixels (resImWidth = resImHeight = 64), and the resulting map is resized back to the image size. This low working resolution explains the coarse appearance of the masks.

$$F(u, v) = \mathscr{F}\{I(x, y)\} = A(u, v) \cdot \exp(j\varphi(u, v)) \quad (2)$$

The log-amplitude spectrum and the spectral residual are given in Equations (3) and (4), where ε is a small constant, h(u, v) is a local averaging filter, and * denotes convolution. This form is a conceptual explanation of the spectral residual; the actual implementation uses the OpenCV function.

$$L(u, v) = \log(A(u, v) + \varepsilon) \quad (3)$$

$$R(u, v) = L(u, v) - h(u, v) * L(u, v) \quad (4)$$

The saliency map is reconstructed from the residual amplitude and the original phase, as in Equation (5). It is then normalized to the range 0–255; in the notebook implementation, the saliency map is converted to uint8 before thresholding.

$$S_0(x, y) = |\mathscr{F}^{-1}\{\exp(R(u, v) + j\varphi(u, v))\}|^2 \quad (5)$$

A binary mask was formed using a threshold T = 25, as in Equation (6), where S(x, y) is the saliency map. A global threshold is a heuristic decision and can produce masks that are too wide or too narrow under different visual conditions. The value T = 25 was fixed a priori and was not tuned on the validation set. Quantitative mask statistics, such as the foreground-area ratio per class, the proportion of empty or near-full masks, and the IoU against manually annotated masks, were not computed in this experiment; mask quality is therefore assessed only qualitatively (Figure 5).

$$M_0(x, y) = 1 \text{ if } S(x, y) \geq T; 0 \text{ if } S(x, y) < T \quad (6)$$

The mask was refined with a closing operation, that is, dilation followed by erosion, using an elliptical structuring element B of size 7 × 7 for two iterations, as in Equation (7), where $\oplus$ denotes dilation and $\ominus$ denotes erosion. Closing can fill small holes and connect disconnected regions, but it does not perform semantic correction of the object.

$$M_{closed} = (M_0 \oplus B) \ominus B \quad (7)$$

The foreground image was obtained by keeping the image pixels inside the mask and replacing the area outside the mask with a white background, as in Equation (8). The configuration used in this experiment is summarized in Table 10.

$$I_{fg}(x, y) = I(x, y) \text{ if } M_{closed}(x, y) = 1; (255, 255, 255) \text{ otherwise} \quad (8)$$

Table 10. Foreground extraction configuration

Component	Actual configuration
Saliency detector	OpenCV StaticSaliencySpectralResidual
Threshold	25
Morphological kernel	Ellipse, 7 × 7
Closing iterations	2
Background outside the mask	White
Images processed	5,140; 0 files missing

Source: Data Processed

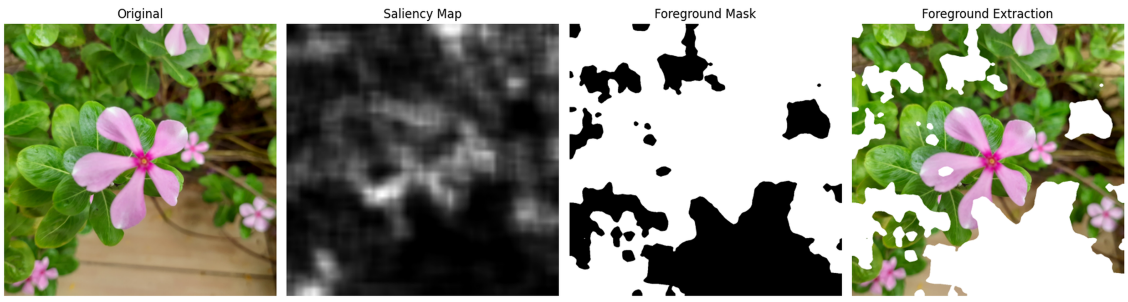


Figure 5. Actual sample from the notebook: original image, saliency map, foreground mask, and foreground extraction result. The image was enlarged and sharpened for readability; the content of the mask was not reconstructed

Figure 5 shows an important limitation: spectral residual saliency highlights visual contrast and structure, not the flower object semantically. In some samples, leaves and background are retained, whereas parts of the object can be cut off or left with holes. This can change the shape of the flower that the model learns.

MobileNetV3Small Model and Experimental Design

MobileNetV3Small pretrained on ImageNet was used with include_top = False. The backbone was frozen and followed by GlobalAveragePooling2D, Dropout 0.2, and Dense(5, softmax). The training configuration is shown in Table 11. Experiment A (baseline) used the original images, and Experiment B used the foreground-extracted images. The model, split, training configuration, and evaluation were kept identical, so the main variable compared is the image preprocessing. Only the frozen-backbone setting was tested; fine-tuning the backbone could change how the network responds to foreground-extracted images and was not evaluated in this study.

Table 11. Model and training configuration

Parameter	Value
Backbone	MobileNetV3Small, ImageNet pretrained
Input	224 × 224 × 3
Backbone trainable	False (frozen)
Classification head	GAP → Dropout(0.2) → Dense(5, softmax)
Optimizer	Adam
Learning rate	0.001
Loss	Sparse categorical cross-entropy
Batch size	8
Maximum epochs	15
EarlyStopping	Monitor val_loss; patience 3; restore_best_weights
Hardware	Google Colab CPU

Source: Data Processed

Evaluation Metrics

Accuracy, precision, recall, and the F1-score were computed as in Equations (9)–(12), where TP_i, FP_i, and FN_i are the true positives, false positives, and false negatives of class i and N is the number of test samples. Macro averages, as in Equation (13), give equal weight to every class; macro-precision and macro-recall are computed in the same way as macro-F1 over K = 5 classes.

$$\text{Accuracy} = (\Sigma_i\, TP_i) / N \tag{9}$$

$$\text{Precision}_i = TP_i / (TP_i + FP_i) \tag{10}$$

$$\text{Recall}_i = TP_i / (TP_i + FN_i) \tag{11}$$

$$F1_i = 2 \times \text{Precision}_i \times \text{Recall}_i / (\text{Precision}_i + \text{Recall}_i) \tag{12}$$

$$\text{Macro-F1} = (1/K)\, \Sigma_{i=1}^{K}\, F1_i \tag{13}$$

The loss function was sparse categorical cross-entropy, as in Equation (14), where $p_{n,yn}$ is the probability that the model assigns to the true class of sample n.

$$L = -(1/N)\, \Sigma_n \log p_{n,yn} \tag{14}$$

Statistical Analysis

Because both models were evaluated on the same 771 test images, the difference between paired predictions was tested with the exact McNemar test. The test uses only the discordant images, where b is the number of images classified correctly by the original model and incorrectly by the foreground model, and c is the reverse. Under the null hypothesis of no difference, the two-sided p-value is obtained from the binomial distribution with n = b + c trials and probability 0.5, as in Equation (15), where C(n, i) is the binomial coefficient. The uncertainty of test accuracy was expressed with exact Clopper–Pearson 95% confidence intervals, and a significance level of 0.05 was used.

$$p = \min\{1,\, 2\, \Sigma_{i=0}^{\min(b,c)}\, C(n, i) \cdot 0.5^n\} \tag{15}$$

RESULTS AND DISCUSSION

EDA Results and Their Relevance to Saliency

The dataset consists of five classes with proportions of 19.46–21.01%, so class imbalance is not extreme. The images are uniform at a resolution of 1024 × 1024, but the EDA showed heterogeneity in brightness, contrast, RGB, background, and blur. Crape Jasmine had a mean brightness of 82.05 and 55 brightness outliers (5.39%). Wedelia had a brightness standard deviation of 25.66 and a mean contrast of 36.97. Madagascar Periwinkle–Wedelia had the closest mean RGB distance, 15.76.

The EDA findings are a reason to test preprocessing, not evidence that the background causes errors. Background blur and object blur must be separated because they have different implications. Global brightness and contrast also include the background, so they do not directly indicate the quality of the flower region.

Foreground Extraction Results

The notebook successfully produced foreground images for all 5,140 images, with no missing files. The visualizations show that the masks are not perfectly consistent: some background areas are retained as foreground, some parts of the flowers may have holes or be removed, and the shape of the object is sometimes altered. Morphological closing can close small gaps, but it cannot determine whether a region is a petal, a leaf, or background.

This indicates that the threshold of 25 involves a trade-off. A low threshold tends to retain a wider salient area, whereas a high threshold can reduce the area considered salient but risks losing parts of the flower. The ideal threshold should be determined through validation, not on the test set.

Performance Comparison

The original model correctly classified 771 of 771 test images, whereas the foreground model correctly classified 768 of 771 and produced three errors (Table 12). The difference in accuracy of −0.3891 percentage points indicates a small decrease, not an improvement. The test loss of the foreground model was higher, indicating that its predictions were probabilistically worse in this evaluation even though accuracy remained high. However, this numerical decrease is based on only three discordant images ($b = 3$, $c = 0$), and the exact McNemar test gave $p = 0.25$. The difference is therefore not statistically significant, and the results are more accurately described as showing no detectable difference between the two conditions. The 95% Clopper–Pearson confidence intervals for accuracy also overlap: 99.52–100% for the original condition and 98.87–99.92% for the foreground condition.

Table 12. Performance comparison on the test set (771 images)

Metric	Original	Foreground	Change (FG − Original)
Accuracy	1.000000 (100%)	0.996109 (99.6109%)	−0.3891 percentage points
Macro-precision	1.000000	0.996061	−0.003939
Macro-recall	1.000000	0.996052	−0.003948
Macro-F1	1.000000	0.996052	−0.003948
Test loss	0.000972	0.006489	+0.005517
Accuracy 95% CI	99.52–100%	98.87–99.92%	Overlapping
Exact McNemar test	–	b = 3, c = 0	p = 0.25 (n.s.)

Source: Data Processed

Error Analysis

The three errors of the foreground model are listed in Table 13. The Hibiscus → Marigold and Marigold → Hibiscus errors indicate an exchange between two classes. Similarity in colour or structure and the loss of detail due to the mask are possible explanations, but they need to be checked through the saliency mask and foreground image of each sample. The Crape Jasmine → Madagascar Periwinkle error also needs to be examined to determine whether the petals or other distinguishing areas remained intact.

Table 13. Misclassified test images in the foreground condition

File	True class	Predicted class
CJ_0036.jpg	Crape Jasmine	Madagascar Periwinkle
H_0346.jpg	Hibiscus	Marigold
M_0821.jpg	Marigold	Hibiscus

Source: Data Processed

The interpretation of errors must be exploratory. Without individual visualizations and subgroup evaluation, it cannot be determined whether the cause of the errors is background removal, visual similarity, or other factors. Hibiscus–Marigold is the third-closest pair in global mean RGB distance (Table 8), which is consistent with, but does not prove, a colour-related confusion between these two classes. Visualizations of the three misclassified images together with their saliency masks were not retained in the current experimental output; they should be reported in follow-up work because they constitute the most direct evidence of the mechanism behind these errors.

Runtime and Trade-offs

Saliency preprocessing took about 537.96 seconds, or 8 minutes 58 seconds (Table 14). Training on the foreground images was 145.27 seconds shorter, but the total foreground preprocessing and training time was 2,284.54 seconds. This is 392.70 seconds (20.76%) longer than baseline training. The recorded inference time for the foreground model does not include saliency extraction for new images; therefore, it is not an end-to-

end inference time. Averaged over 5,140 images, saliency extraction required about 0.105 s per image, which would be added to end-to-end inference for every new image.

The recorded test-set inference times (40.99 s and 15.15 s, or about 53.2 ms and 19.7 ms per image) were each measured once for the same architecture and input size; the difference most likely reflects caching or warm-up effects rather than a real difference in model speed and should not be interpreted as evidence that the foreground model is faster. A controlled benchmark with warm-up runs and repeated timing is required for a valid comparison.

Table 14. Runtime comparison

Component	Original	Foreground
Saliency preprocessing	0 s	537.9628 s
Training	1,891.8450 s	1,746.5773 s
Preprocessing + training	1,891.8450 s	2,284.5401 s
Inference on the test set	40.9880 s	15.1525 s

Source: Data Processed

General Discussion

The results show that reducing the background does not automatically improve classification. In this dataset, the background may contain contextual information that is not entirely harmful, which is consistent with the observation that models can exploit both foreground and background cues (Moayeri et al., 2022). At the same time, saliency masks can remove important parts of the flower. Spectral residual saliency is a bottom-up approach (Hou & Zhang, 2007) and was not designed specifically for flower segmentation; learned salient object detectors or plant-specific segmentation models (Ma et al., 2023; Wang et al., 2022) may produce more faithful masks, but at a higher computational cost. From a methodological point of view, the experiment yields an important negative result: foreground extraction maintained high performance but did not improve it on the test split. Negative results are informative because they show that preprocessing methods need to be evaluated together with mask quality and computational cost rather than simply assumed to be beneficial (KC et al., 2021).

These findings are consistent with previous studies indicating that the benefit of background removal is not universal. KC et al. (2021) applied segmentation-based background removal to improve CNN accuracy on in-situ plant images with cluttered backgrounds, but the benefit depended on the dataset and on the quality of segmentation. Moayeri et al. (2022) showed that classifiers draw on both foreground and background information, so suppressing the background can remove useful context as well as noise. Unlike segmentation methods designed for plant parts (Guo et al., 2023; Ma et al., 2023), the present study used a training-free, bottom-up saliency method (Hou & Zhang, 2007) whose masks are coarse and not object specific, which may explain why no benefit was observed.

The large variation in the performance of salient object detection techniques across images reported by Ahmed et al., (2022) is also consistent with the inconsistent masks

observed in Figure 5. With respect to performance level, the baseline accuracy obtained here (100%) is higher than that reported for other medicinal flower datasets, such as 97.64% for Chinese medicinal flowers with an optimized ResNet101 (Huang & Xu, 2023). This difference should not be interpreted as evidence of the superiority of MobileNetV3Small, because the datasets, the number of classes, and the evaluation protocols differ.

Rather, it suggests that the five-class PhulMedic task under a random image-level split is relatively easy for this backbone, possibly further simplified by near-duplicate images. Under such ceiling conditions, a preprocessing method can only appear neutral or harmful, never beneficial, so the absence of improvement in this study reflects the evaluation design as much as the method itself. More informative designs would include repeated stratified k-fold cross-validation with confidence intervals, group-aware splits after near-duplicate removal, cross-condition or background-swap robustness tests, or test images from an independent source.

Validity and Limitations

1. The baseline achieved 100% accuracy; the results need to be audited because this may indicate an easy dataset or possible leakage.
2. The notebook detected duplicate pairs in Crape Jasmine. It must be confirmed that they were not distributed across subsets; if they were, deduplication or a group split should be applied and the experiment repeated
3. The study used only one split and one run; seed variation and confidence intervals are not yet available
4. No ground-truth masks are available to compute IoU or Dice for foreground quality.
5. The visual EDA has not yet been turned into a controlled robustness test for background, brightness, contrast, or blur
6. The inference runtime of the foreground model does not yet include saliency preprocessing for new images.
7. The numerical results are limited to the dataset, preprocessing, threshold, backbone, and experimental configuration used.
8. The difference between the two conditions (three errors versus none) is not statistically significant (exact McNemar $p = 0.25$).
9. Only a frozen backbone was evaluated; the effect of foreground extraction under fine-tuning was not tested

CONCLUSION

This study evaluated whether saliency-guided foreground extraction based on StaticSaliencySpectralResidual improves MobileNetV3Small classification of five medicinal plant flower classes in the PhulMedic dataset. The foreground condition achieved 99.61% accuracy and macro-F1 on 771 test images, compared with 100% for the original images; this difference was not statistically significant (exact McNemar $p = 0.25$), indicating no detectable difference between the conditions.

Foreground extraction added about 20.76% computational overhead relative to baseline training without providing an accuracy benefit. The key methodological lesson is that, with a saturated baseline and a single split, the experiment cannot evaluate whether foreground extraction helps. Future work should audit near-duplicate leakage, use group-aware repeated cross-validation, quantify mask quality, and test fine-tuned backbones. These conclusions are limited to the PhulMedic dataset, the frozen MobileNetV3Small backbone, and the preprocessing configuration used.

FUNDING

The author received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors for this research or for the publication of this article. The experiments were conducted using the publicly available PhulMedic dataset and freely accessible computing resources (Google Colab CPU).

CONFLICT OF INTEREST

The author declares that there are no conflicts of interest regarding the publication of this article. The research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The author has no affiliation with the creators of the PhulMedic dataset or with the developers of the software libraries used in this study.

AUTHOR CONTRIBUTION

Viandre Chika Baratama: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Visualization, Writing – original draft, and Writing – review & editing. The author has read and approved the final version of the manuscript and takes full responsibility for its content.

GENERATIVE ARTIFICIAL INTELLIGENCE USAGE STATEMENT

During the preparation of this manuscript, the author used generative artificial intelligence (AI) tools to improve language clarity and grammatical structure and to help search for, verify. The AI tools were not used to generate research data, conduct the experiments, perform the statistical analysis, or interpret the results. After using these tools, the author reviewed and edited all content, verified every cited reference against its original source, and takes full responsibility for the integrity and accuracy of the published article. This use complies with applicable ethical guidelines on the disclosure of generative AI in scientific writing.

REFERENCES

Abbas, T., Razzaq, A., Zia, M. A., Mumtaz, I., Saleem, M. A., Akbar, W., Khan, M. A., Akhtar, G., & Shivachi, C. S. (2022). Deep Neural Networks for Automatic Flower Species Localization and Recognition. *Computational Intelligence and Neuroscience*, *2022*, 1–9. https://doi.org/10.1155/2022/9359353

Ahmed, K., Gad, M. A., & Aboutabl, A. E. (2022). Performance evaluation of salient object detection techniques. *Multimedia Tools and Applications*, *81*(15), 21741–21777. https://doi.org/10.1007/s11042-022-12567-y

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*(1), 53. https://doi.org/10.1186/s40537-021-00444-8

Borji, A., Cheng, M.-M., Hou, Q., Jiang, H., & Li, J. (2019). Salient object detection: A survey. *Computational Visual Media*, *5*(2), 117–150. https://doi.org/10.1007/s41095-019-0149-9

Cao, S., & Song, B. (2021). Visual attentional-driven deep learning method for flower recognition. *Mathematical Biosciences and Engineering*, *18*(3), 1981–1991. https://doi.org/10.3934/mbe.2021103

Chen, Y., Huang, Y., Zhang, Z., Wang, Z., Liu, B., Liu, C., Huang, C., Dong, S., Pu, X., Wan, F., Qiao, X., & Qian, W. (2023). Plant image recognition with deep learning: A review. *Computers and Electronics in Agriculture*, *212*, 108072. https://doi.org/10.1016/j.compag.2023.108072

Fei, Y., Li, Z., Zhu, T., & Ni, C. (2023). A Lightweight Attention-Based Convolutional Neural Networks for Fresh-Cut Flower Classification. *IEEE Access*, *11*, 17283–17293. https://doi.org/10.1109/ACCESS.2023.3244386

Guo, Y., Sun, K., Luo, G., & Wang, M. (2023). Unsupervised leaf segmentation in complex backgrounds using mutual information minimization. *Journal of Intelligent & Fuzzy Systems*, *45*(6), 10127–10139. https://doi.org/10.3233/JIFS-232696

Hou, X., & Zhang, L. (2007). Saliency Detection: A Spectral Residual Approach. *2007 IEEE Conference on Computer Vision and Pattern Recognition*, 1–8. https://doi.org/10.1109/CVPR.2007.383267

Howard, A., Sandler, M., Chen, B., Wang, W., Chen, L.-C., Tan, M., Chu, G., Vasudevan, V., Zhu, Y., Pang, R., Adam, H., & Le, Q. (2019). Searching for MobileNetV3. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 1314–1324. https://doi.org/10.1109/ICCV.2019.00140

Huang, M.-L., & Xu, Y.-X. (2023). Image classification of Chinese medicinal flowers based on convolutional neural network. *Mathematical Biosciences and Engineering*, *20*(8), 14978–14994. https://doi.org/10.3934/mbe.2023671

Huang, M.-L., Xu, Y.-X., & Liao, Y.-C. (2021). Image dataset on the Chinese medicinal blossoms for classification through convolutional neural network. *Data in Brief*, *39*, 107655. https://doi.org/10.1016/j.dib.2021.107655

KC, K., Yin, Z., Li, D., & Wu, Z. (2021). Impacts of Background Removal on

Convolutional Neural Networks for Plant Disease Classification In-Situ. *Agriculture*, *11*(9), 827. https://doi.org/10.3390/agriculture11090827

Khan, A. T., Jensen, S. M., Khan, A. R., & Li, S. (2023). Plant disease detection model for edge computing devices. *Frontiers in Plant Science*, *14*. https://doi.org/10.3389/FPLS.2023.1308528/FULL

Ma, R., Fuentes, A., Yoon, S., Lee, W. Y., Kim, S. C., Kim, H., & Park, D. S. (2023). Local refinement mechanism for improved plant leaf segmentation in cluttered backgrounds. *Frontiers in Plant Science*, *14*. https://doi.org/10.3389/fpls.2023.1211075

Malik, O. A., Ismail, N., Hussein, B. R., & Yahya, U. (2022). Automated Real-Time Identification of Medicinal Plants Species in Natural Environment Using Deep Learning Models—A Case Study from Borneo Region. *Plants*, *11*(15), 1952. https://doi.org/10.3390/plants11151952

Moayeri, M., Pope, P., Balaji, Y., & Feizi, S. (2022). A Comprehensive Study of Image Classification Model Sensitivity to Foregrounds, Backgrounds, and Visual Attributes. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 19087–19097. https://doi.org/10.1109/CVPR52688.2022.01850

PhulMedic: A Bangladeshi Medicinal Flower Dataset (Version 2), Mendeley Data (2025). https://doi.org/10.17632/92n7gtrkhx.2

Rajkomar, G., & Pudaruth, S. (2023). A Mobile App for the Identification of Flowers Using Deep Learning. *International Journal of Advanced Computer Science and Applications*, *14*(5). https://doi.org/10.14569/IJACSA.2023.0140508

Roopashree, S., & Anitha, J. (2021). DeepHerb: A Vision Based System for Medicinal Plants Using Xception Features. *IEEE Access*, *9*, 135927–135941. https://doi.org/10.1109/ACCESS.2021.3116207

Wang, W., Lai, Q., Fu, H., Shen, J., Ling, H., & Yang, R. (2022). Salient Object Detection in the Deep Learning Era: An In-Depth Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(6), 3239–3259. https://doi.org/10.1109/TPAMI.2021.3051099

Wu, R., Feng, M., Guan, W., Wang, D., Lu, H., & Ding, E. (2019). A Mutual Learning Method for Salient Object Detection With Intertwined Multi-Supervision. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8150–8159. https://doi.org/10.1109/CVPR.2019.00834

Zeng, Z., Huang, C., Zhu, W., Wen, Z., & Yuan, X. (2023). Flower image classification based on an improved lightweight neural network with multi-scale feature fusion and attention mechanism. *Mathematical Biosciences and Engineering*, *20*(8), 13900–13920. https://doi.org/10.3934/mbe.2023619