# Journal of Informatics and Web Engineering

Vol. 3 No. 2 (June 2024) eISSN: 2821-370X

# Facial Skin Analysis in Malaysians using YOLOv5: A Deep Learning Perspective

**Ying Huey Gan[1], Shih Yin Ooi[2]*, Ying Han Pang[2], Yi Hong Tay[3] and Quan Fong Yeo[2]**

[1] Public Mutual Berhad, Menara Public Bank 2, No.78, Jalan Raja Chulan, 50200 Kuala Lumpur, Malaysia
[2] Faculty of Information Science and Technology, Multimedia University, Jalan Ayer Keroh Lama, 75450 Melaka, Malaysia
[3] 365 Production, B1-13A-01, Soho Suite @KLCC, 20, Jalan Perak, 50450 Kuala Lumpur, Malaysia
**corresponding author: (syooi@mmu.edu.my, ORCiD: 0000-0002-3024-1011)*

*Abstract* - Nowadays, people are more concerned about their skin conditions and are more willing to spend money and time on facial care routines. The beauty sector market is increasing, and more skin type readers are being created to help people determine their skin type. While various skin type readers are in the market, each is invented and tested abroad. Those skin type readers in the beauty market are not applied well on Malaysian skin. Therefore, this paper proposes a facial skin analysis system tailored primarily for Malaysian skin. This paper integrated object detection and deep learning algorithms in developing skin-type readers. A unique dataset consisting solely of facial images of Malaysian skin was created from scratch for the model. Additionally, You Only Look Once version 5 (YOLOv5) is employed to detect users' facial skin conditions, such as acne, pigment, enlarged pores, uneven skin, blackheads, etc. Then, based on the detected skin conditions, it further classifies the user's skin type into the normal, oily, sensitive, or dry groups.

*Keywords— Skin Type Classification, Image Processing, Object Detection, Deep Learning, YOLOv5*

Received: 14 June 2023; Accepted: 23 December 2023; Published: 16 June 2024

## I. INTRODUCTION

Skin can be considered as the human body's largest organ. It is crucial not only for health but also for appearance. Good skin care is vital at any age. In the first few decades of life, the skin has a substantial supply of elastin and collagen, but it gradually diminishes. In addition, daily lifestyle can also directly affect the appearance of human skin. For example, high consumption of fats and sugars can result in blemishes and acne. Hence, everyone is concerned with the condition of their facial skin. More people invest in skin care products to care for their facial skin. First, they must know their skin type to take appropriate steps. Many people do not know how to diagnose their skin type because they lack knowledge of their skin conditions, the characteristics of different skin types, and the proper skin analysis tool. This situation can lead them to apply the wrong skin care products to their facial skin, causing problems such as pimples, wrinkles, facial redness, and dry skin.

To overcome this issue, various types of skin analysis tools exist to help users determine their skin type. A skin analysis tool can diagnose skin conditions and provide skin types in a second with just a snap. Furthermore, with the help of Artificial Intelligence (AI) technology, the results will be more relevant and accurate due to the algorithm used in the tools. It can analyze the skin information through camera detection and, based on its knowledge of various conditions, give the most likely match to the user's face. With the usage of this proposed system, it can also

Journal of Informatics and Web Engineering
https://doi.org/10.33093/jiwe.2024.3.2.1

Published by MMU Press. URL: https://journals.mmupress.com/jiwe

significantly avoid skin type identification errors due to the lack of skincare knowledge. Besides, some skin analysis tools can recommend to users the skin care products that work best for their skin type. It greatly benefits the user as it minimizes the possibility of users choosing the incorrect products to apply to their facial skin. A skin analysis tool can improve human skin health.

However, no skin analysis tools have been invented, primarily for Malaysians' facial skin. Most skin analysis tools are developed for foreign skin, and those may not be applied well to Malaysians. Therefore, this paper focuses on analyzing the facial skin of Malaysians with YOLOv5 technology. A skin type reader that is tolerant of the Malaysians' skin is introduced in this research. It can assist Malaysians in detecting their skin conditions and classifying their skin types correctly and accurately. In this case, the system is developed based on the Malaysians' skin type and tested using Malaysians' skin conditions and characteristics. Facial skin images of Malaysians are collected as the datasets used to train the YOLOv5 model.

## II. LITERATURE REVIEW

Machine learning and deep learning algorithms have shown robust classification results in skin quality examination. This section reviews and tests a few machine learning and deep learning algorithms to analyze facial skin. Research on facial skin analysis has been conducted for many years, but most of them are focused on skin diseases analysis [1,2,3,4]. Although there are also studies on skin types or skin condition analysis [5,6], they are mainly developed on facial skin of foreigners. To date, there are no research has been performed on skin type classification primarily for Malaysians.

### *A. Traditional Machine Learning Approach*

In this section, various type of facial skin analysis using traditional machine learning algorithms are reviewed.

#### *i. Fuzzy C-Means*

Fuzzy c-means (FCM) is one of the most widely used fuzzy clustering algorithms developed by [7] and improved by [8]. It refers to a clustering algorithm that extends the traditional "k-means" algorithm. The FCM algorithm aims to divide a finite collection of *n* elements, represented as Equation (1)

$$X = \{x1, \ldots, xn\} \tag{1}$$

into c fuzzy clusters based on a specified criterion. When applied to a finite set of data, the algorithm produces a set of C cluster centers denoted as Equation (2)

$$C = \{c1, \ldots, cc\} \tag{2}$$

along with a partition matrix, as Equation (3)

$$W = wi, j \in [0,1],\ i = 1, \ldots, n,\ j = 1, \ldots, c \tag{3}$$

Each element *wij* in the partition matrix indicates the degree to which the element xi belongs to the cluster *cj*. The goal of FCM is to minimize an objective function; refer to Equation (4) – (5):

$$J(W,C) = \sum_{i=1}^{n}\sum_{j=1}^{c} w_{ij}^{m} \parallel xi - cj \parallel^2 \tag{4}$$

where,

$$w_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{\parallel x_i - c_j \parallel}{\parallel x_i - c_k \parallel}\right)^{\frac{2}{m-1}}} \tag{5}$$

The research group [9] adopted the basis of FCM with some modifications to analyze facial skin conditions. They proposed a classification system for facial skin conditions based on two-stage Fuzzy c-means. The system consisted of three steps: image capture, feature extraction, and pattern analysis. They utilized a BS-88E Beauty Scope sensor to capture images and focused on the central region of interest (ROI) in the grey part of the image. Feature extraction involved computing contrast, entropy, inverse difference moment (IDM), and deviation information from the ROI and calculating four wavelet parameters to improve time-frequency information [9]. The pattern analysis stage employed a two-stage FCM algorithm to analyze patterns in this system. The first stage clustered the skin images into oily and uncertain groups based on wavelet parameters. In contrast, the second stage further classified the dry and neutral skin groups using texture features like contrast, entropy, deviation values, and IDM [9]. The proposed system achieved a classification accuracy of 93% for facial skin conditions.

*ii. Fuzzy Modelling*

A fuzzy model is a mathematical representation of a system or problem that incorporates fuzzy logic principles to handle uncertainty, imprecision, and vagueness developed by [10]. It can recognize, represent, manipulate, interpret, and utilize data and information lacking certainty and imprecise. The fuzzy model can be defined in terms of sigmoid function. For example, the standard logistic function is defined as in Equation (6),

$$S(x) = \frac{1}{1+e^{-x}} \tag{6}$$

which has the following symmetry property, as in Equation (7) – (8):

$$S(x) + S(-x) = 1 \tag{7}$$

$$(S(x) + S(-x)) \cdot (S(y)) \cdot (S(z) + S(-z)) = 1 \tag{8}$$

The fuzzy model is adopted by [11] for facial skin analysis. They developed a smartphone-based application to examine facial skin quality. The application compressed the captured image and converted it into a grayscale image. Then, it computed texture features such as contrast, IDM, and entropy through the grey-level co-occurrence matrix (GLCM). Skin type such as dry, neutral, or oily was classified using a fuzzy model based on contrast and IDM [11]. Additional texture features such as horizontal texture (LH), vertical texture (HL), and diagonal texture (HH) obtained through first-order Haar DWT were used to detect oily skin because oily skin has a complex texture. Moreover, to minimize the feature dimensions and speed up the detection of skin quality, they employed principal component analysis (PCA) to distinguish texture features. As a result, the proposed fuzzy-model-based skin quality detection application achieved 96.29% and 93.21% accuracy in PCA and GLCM, respectively [11].

*iii. Support Vector Machine*

Support Vector Machine (SVM) is a supervised machine learning algorithm invented by [12] to tackle classification and regression tasks. Computing an SVM classifier is equivalent to a minimized form of expression, as in Equation (9):

$$\left[\frac{1}{n}\sum_{i=1}^{n} max\left(0,1-y_i(w^{\intercal}x_i-b)\right)\right]+\lambda \parallel w \parallel^2 \tag{9}$$

SVM algorithm is adopted to classify skin quality through a skincare recommendation system based on SVM [13]. The system consists of two sections: acne detection and skin type classification. Facial images were captured using a Logitech C310 camera, and the face region in the image was then detected using the Open-Source Computer Visual Library (OpenCV) face recognition model. Specific areas which prone to acne and oil were selected as ROI images. The ROI images were converted to grayscale and underwent a two-level Haar discrete wavelet transform (DWT) to obtain seven sub-bands. Parameters derived from the sub-bands were used for the first level of skin type classification. In the second level of skin type classification, texture features were calculated using the ROI image transformed by the Haar wavelet, where a co-occurrence matrix computed cumulative numbers based on pixel angle and distance. Contrast, IDM, and entropy were derived from the co-occurrence matrix and used in the second-level skin type classification process [13]. Also, SVM was applied to differentiate between oily and non-oily skin based on the first type of features and to differentiate dry and neutral skin from non-oily skin based on the second type of features [13]. Undeniably, SVM excels in classifying from explicit features, and training classification models only need a modest number of samples.

*B. Deep Learning Approach*

*i. Convolutional Neural Network*

A convolutional Neural Network (CNN) is an artificial neural network that processes and analyzes visual data such as images or videos, introduced by [14]. It consists of several layers: convolutional, activation, pooling, fully connected, dropout, and normalization. When putting them into a specific order, these layers allow CNN to learn hierarchical representations of input data and perform tasks like image classification, object detection, and image generation. In convolutional layer, it is formulated by the Equation (10),

$$W_{out} = \frac{W - F + 2P}{S} + 1 \tag{10}$$

If considering an input with dimensions W x W x D and $D_{out}$ kernels with a spatial size of F, a stride of S, and padding of P, the size of the output volume can be determined using the formula mentioned above. For the pooling layer, the equation is defined in Equation (11),

$$W_{out} = \frac{W - F}{S} + 1 \tag{11}$$

Considering an activation map with dimensions W x W x D, a pooling kernel with a spatial size of F, and an S stride, the output volume size can be determined using the abovementioned formula.

Next, the non-linearity layers. Following the linear operation of convolution, non-linear layers are typically added directly after the convolutional layer to introduce non-linearity to the activation map. This step is necessary because images exhibit non-linear characteristics, and non-linearity layers help capture and represent such complexities effectively. Firstly, the sigmoid function used within a layer of CNN is formulated by Equation (12),

$$\sigma(x) = 1 / (1 + e^{-x}) \tag{12}$$

Besides, the activation function, ReLU, sets all negative input values to zero and leaves positive values unchanged. It is represented as in Equation (13):

$$f(x) = max\ (0, x) \tag{13}$$

Both functions are frequently used in CNN after the convolutional layers to introduce non-linearity.

[15] have studied an application for skin type classification and skin care product recommendation based on a Convolutional Neural Network. They created a dataset of over eighty skin images categorized as dry or oily skin and trained the CNN model using this dataset. The model achieved an accuracy of 85% in classifying skin types from facial images [15]. The CNN architecture included two convolutional layers followed by max pooling layers. The input image size was 100 × 100 pixels with RGB spectrum. The Conv+MaxPool layer extracted in-depth features, resulting in a data size of 23 × 23 × 32. The flattened layer converted the data to a single dimension, and the subsequent dense layers provided the skin type results, distinguishing between dry and oily skin using a Sigmoid activation function. In other words, the skin is considered dry when the Sigmoid output is closer to one; the skin is thought to be oily when the Sigmoid output is closer to zero [15].

*ii. Residual Network*

A residual network, or ResNet, is a deep learning architecture introduced by [16] for computer vision applications. It was developed to address the problem of vanishing gradients in intense neural networks. ResNet is formulated as the Equation (14),

$$y = f(x) + x \tag{14}$$

where x represents the input to the residual unit, and f(x) corresponds to a series of convolutional layers, typically consisting of two layers or three in the case of bottleneck layers. To ensure clarity, use XL to refer to all inputs to layer L and XL+1 to represent the corresponding output, which serves as the input for the next layer, L+1. With this notation in mind, the equation for the residual unit can be written as in Equation (15) – (16):

$$YL = f(xL) + XL \tag{15}$$

$$XL{+}1 = RELU(YL) \tag{16}$$

In this equation, the original input is added to the output of the layers without any modification. Thus, for any layer L+N, the network equation is expressed as in Equation (17):

$$XL{+}N = XL + f(XL) + f(XL{+}1) + f(XL{+}2) + \ldots. + f(XL{+}N{-}1) \tag{17}$$

[17] adopted the basis of ResNet with some modifications to analyze skin diseases. They introduced a skin disease classification based on a thirty-four-layer ResNet. Around 587 images of 11 skin conditions and rashes were collected and labeled using AWS SageMaker Ground Truth [17]. The ResNet algorithm was trained using MXNet, with 30 epochs, a learning rate 0.1, and an image shape of 244 x 244 [17]. Batch normalization was employed for improved performance.

Furthermore, the ResNet model was compared to the industry standard Google AutoML Vision model, and the ResNet model outperformed in terms of precision, recall, and average precision for skin disease detection and classification [17]. To prove this, the ResNet model had a precision of 0.925, a recall of 0.681, and an average precision of 0.917. In contrast, the Google AutoML model had a precision of 0.8982, a recall of 0.634, and an average precision of 0.887. Thus, ResNet's reformulated layers with residual functions enable it to produce results with higher precision and fewer training errors.

*iii. YOLOv5's Novel Object Recognition Algorithm*

YOLOv5 is a state-of-the-art object detection model developed by [18]. It is an evolution of the YOLOv5 model series, known for its real-time object detection capabilities. The YOLOv5 algorithm uses a specific loss function called the YOLO loss function. The YOLO loss function combines multiple components to calculate the overall loss during training. It includes the following components, bounding box regression loss, classification loss, and confidence loss.

The bounding box regression loss is defined as in Equation (18),

$$l_{box} = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{i,j}^{ojb} bj(2 - w_i \times h_i)\left[\left(x_i - \hat{x}_i^j\right)^2 + \left(y_i - \hat{y}_i^j\right)^2 + \left(w_i - \widehat{w}_i^j\right)^2 + \left(h_i - \hat{h}_i^j\right)^2\right] \tag{18}$$

The classification loss is defined as in Equation (19),

$$l_{cls} = \lambda_{class} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{i,j}^{obj} \sum_{C \in classes} p_i(c) log\left(\widehat{p_l}(c)\right) \tag{19}$$

The confidence loss is defined in Equation (20),

$$l_{obj} = \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{i,j}^{obj} (c_i - \hat{c}_l)^2 + \lambda_{obj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{i,j}^{obj} (c_i - \hat{c}_l)^2 \tag{20}$$

The total YOLOv5 loss combines these individual losses with appropriate weighting applied to each component, refer to Equation (21).

$$Loss = Lbox + Lcls + Lobj \tag{21}$$

DermaGenics using YOLOv5 to identify melanoma skin cancer is made available to the web [19]. Users can upload photos of skin lesions, and the YOLOv5 model analyzes the images to determine if the lesion is malignant or benign. The system uses novel image processing algorithms to segment the lesion, analyze its shape, size, and texture, and classify it as melanoma or a malignant lesion of normal skin based on derived feature characteristics. The YOLOv5 model used in the application is YOLOv5s, a lightweight version with a size of 14MB. The dataset was divided into training, validation, and testing sets to ensure the model's performance. Finally, the average accuracy of the YOLOv5 model in identifying melanoma and non-melanoma reached 89%, the average accuracy of identifying melanoma reached 93%, and the average accuracy of identifying non-melanoma reached 85% [19].

## III. RESEARCH METHODOLOGY

This section presents how the YOLOv5 model is processed and trained and how the skin type reader analyzes facial skin.

### *A. YOLOv5 Model*

The proposed YOLOv5 [20-22] model comprises five critical processes: data collection, data preparation, image pre-processing, facial skin feature extraction, and skin conditions detection. Figure 1 shows the overall process flow of the YOLOv5 model and depicts how each process works.

#### *i. Data Collection*

Sufficient data must be collected to yield a good trained YOLOv5 model. In this study, the Malaysian skin types were collected from a pool of Multimedia University (MMU) students and communities. Instead of using a fixed studio background photo, the volunteers were requested to submit their facial skin photos to imitate real-life challenges. There are pros and cons to this approach. The advantage is that it mimics real-life challenges, but the disadvantage is that it poses more challenges in machine learning tasks due to diverse backgrounds, illumination, etc. On top of requesting photos, the volunteers were asked to answer some questions through a Google Form about their skin condition. This approach ensures their photos are labeled with the correct class of skin types. The Google Form is divided into six sections, with section 1 asking respondents about their gender and age and sections 2, 3, 4, and 5 asking for details about their skin condition, for example, whether their skin is oily, dry, sensitive, resistant, wrinkle, tight, pigmented or non-pigmented. The last section allows them to upload images of their front, left, and right angles of faces. Guidelines on capturing facial skin are established in Google form to ensure that facial skin images are standardized under similar conditions, allowing for more accurate training of the YOLOv5 model. The guideline includes taking photos in natural light, tying up your hair, removing makeup and glasses, avoiding visual distractions, etc.

Due to factors such as climate, lifestyle, genetics, etc., Malaysian's facial skin differs greatly from that of other nations, necessitating the usage of a local dataset. Firstly, Malaysia has a tropical climate, which is hot and humid all year. High humidity and temperature lead the skin to generate extra oil, making it more susceptible to acne and outbreaks. The humidity can also affect the skin's moisture balance, leading to dehydration or increased water loss. In addition, hot and humid weather can potentially exacerbate redness or flushing in individuals with sensitive or reactive skin. Besides, the intense sunlight and ultraviolet radiation in Malaysia can contribute to uneven skin tones, the development of dark spots, and the exacerbation of pigmentation issues.

#### *ii. Data Preparation*

In the data preparation process, the Malaysians' facial skin images collected in the previous process are prepared to make them into a dataset. The data preparation process involves using LabelImg, a tool for graphically labeling images. It is used to draw bounding boxes around facial skin features such as acne, blackheads, uneven skin, etc., and label them with the corresponding class name. The annotation data was exported into YOLO format, and then, the labeled images and their annotation data were composed into a dataset. After that, a Python program is executed to split the dataset into two folders: the training and validation datasets.

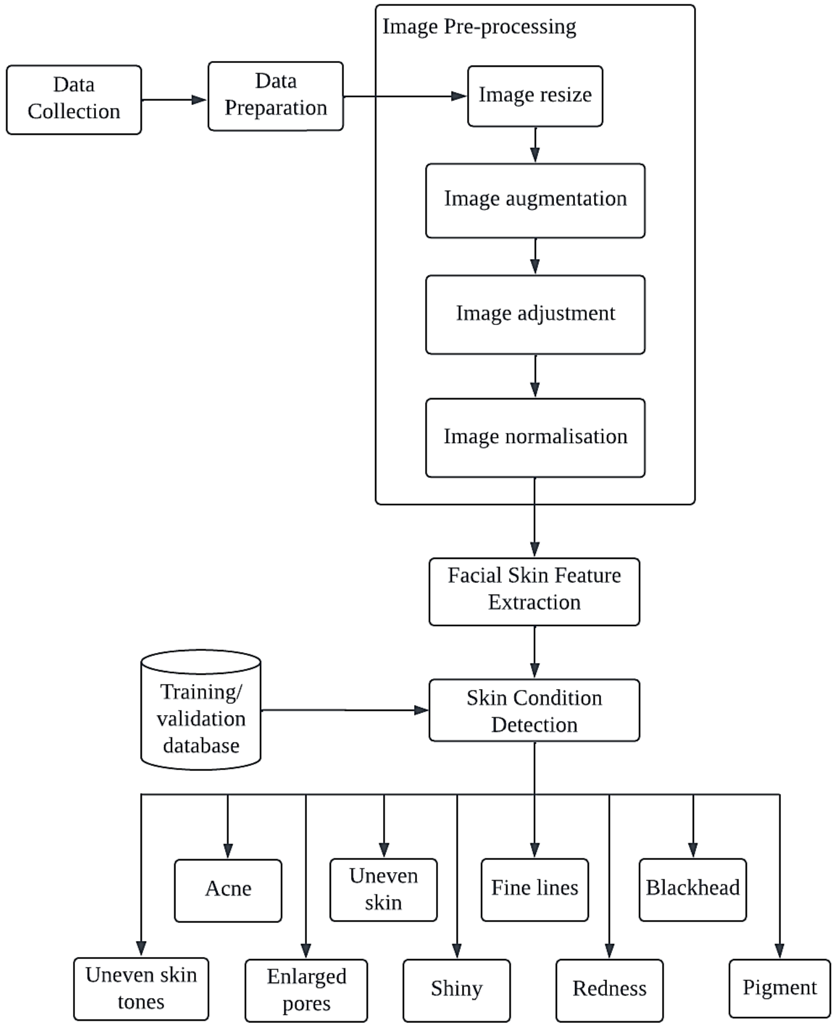


Figure 1. Block Diagram of YOLOv5 Model

*iii. Image Pre-processing and Facial Skin Feature Extraction*

Next, the datasets are passed into the YOLOv5 model. In YOLOv5, there is an image pre-processing process. The image in the datasets is resized because the YOLOv5 model is trained faster on smaller images to ensure that it receives consistent input. Then, data augmentation, such as rotating, flipping, and scaling, is applied to the image to enhance the accuracy and prevent overfitting of the model. After that, adjusting the brightness and contrast of the image improves the perceptibility of the image and makes skin conditions distinguishable. Lastly, normalizing the pixel value of the image to a specific range helps improve the model's performance by ensuring that all images have similar ranges of values. At this point, an enhanced facial skin image is generated.

After image pre-processing, facial skin features are extracted to eliminate the redundant data. For example, facial skin features include acne, uneven skin, fine lines, blackheads, uneven skin tones, enlarged pores, shine, redness, and pigment. The nine facial skin conditions were extracted because the skin conditions of each facial skin image in the dataset passed into the YOLOv5 model were annotated with bounding boxes and labelled with the corresponding class name. When the dataset is transferred to the YOLOv5 model, it takes the annotated images as input and then, learns to recognize the features and extract them. In this case, the YOLOv5 model uses the layers of its convolutional neural network (CNN) to extract features at different scales and levels of abstraction.

*iv. Skin Conditions Detection*

Next, the YOLOv5 model was trained using the training dataset. The YOLOv5 model learns the facial skin features of the images in the training dataset to have the capability to diagnose skin conditions. The validation dataset was also used to test whether the trained YOLOv5 model was effective and unbiased by evaluating the model's performance on previously unseen data. After training the YOLOv5 model, an image can be inputted into the model to verify the result and accuracy of skin conditions detection. The trained YOLOv5 model will provide the skin conditions detection result where the facial features of input facial skin images are bounded with boxes and labeled with their class names.

*B. Facial Skin Analysis*

The proposed skin type reader comprises five significant phases: image acquisition, image pre-processing and feature extraction, skin conditions detection, skin type classification, and skincare product recommendation. Figure 2 shows the skin type reader's overall program flow and depicts how each phase works.

*i. Image Acquisition*

The image reader scheme starts by requesting a facial skin image from the user. The image should be qualified to proceed to the next phase, the pre-processing of the facial skin image. Otherwise, the user must recapture the image. The requirement for a qualified image is that the image is not blurry, the content of the image taken is facial skin, taken in natural light, and the face is free of makeup and glasses.

*ii. Image Pre-processing and Feature Extraction*

The captured image is transferred to the YOLOv5 model for facial skin analysis. Before that, the captured image underwent the pre-processing and feature extraction process of the YOLOv5 model. The pre-processing process comprises image resizing, image augmentation such as rotation, flipping and scaling, brightness and contrast adjustment, and image pixel value normalization, as shown in Figure 3.

*iii. Skin Conditions Detection, Skin Type Classification, Skincare Product Recommendation*

After image pre-processing and feature extraction, the skin conditions of the input image are detected using a trained YOLOv5 model. If the skin conditions are successfully detected, the skin type reader will move into the next phase, which is skin type classification. Otherwise, it will take the user back to the beginning of the facial skin analysis, inputting the facial skin image again.

Next, the skin type of the image is classified according to the skin conditions detected in the previous phase. If the skin type is successfully identified, the result will be displayed to the user; otherwise, the user will be redirected to the start of the facial skin analysis.

A skin type is considered successfully determined when the system provides user with skin type results of dry, oily, normal or sensitive, where user receives their particular skin type at the “SKIN SUMMARY” section of the “Skin Analysis” page. A skin type classification is considered failed when the system is unable to provide skin type results to the user and the user receives an error message at the “SKIN SUMMARY” section requiring the user to reupload the facial skin image.

After that, the classified skin type result generated in the previous phase is transferred to the skincare product recommendation phase to produce products suitable for the user's skin type. For example, face cleanser, toner, moisturizing gel, sunscreen, etc. At the end of the system, data such as users' skin conditions, skin type, and suitable skincare products will be output to the user.

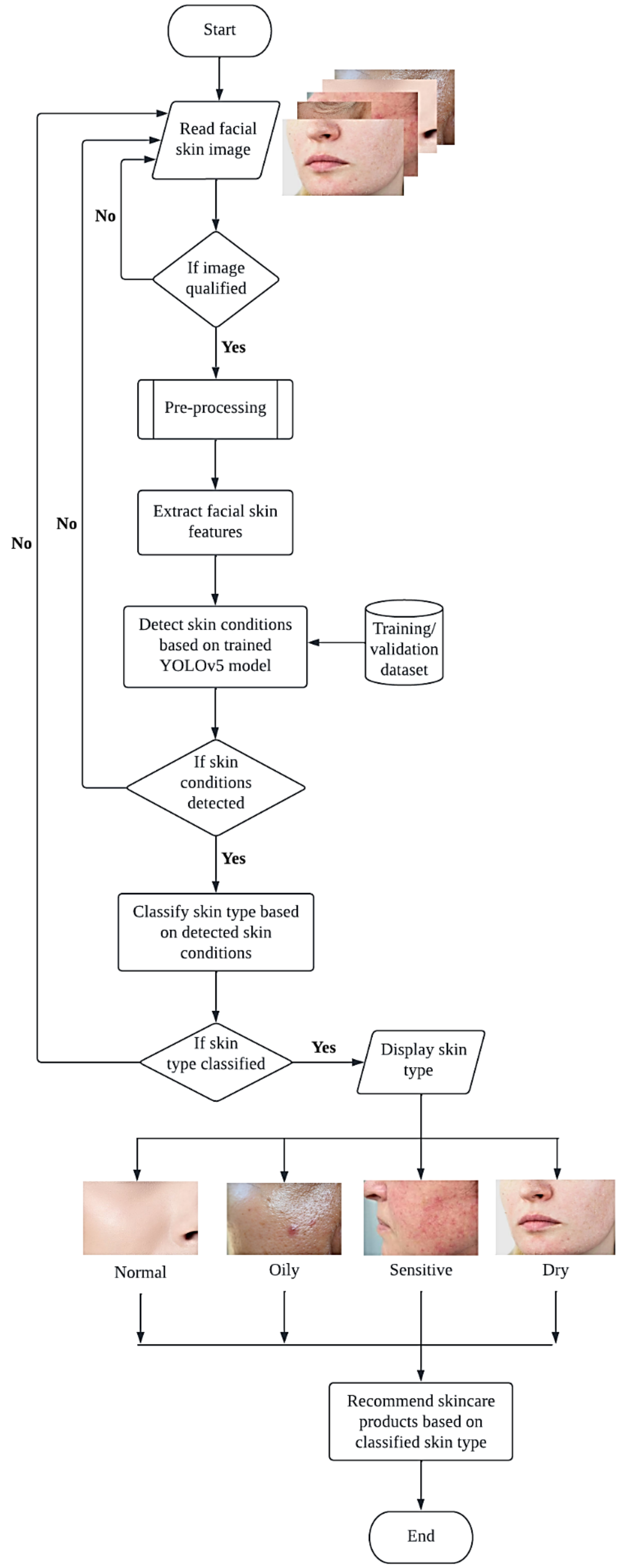


Figure 2. Flowchart of Skin Type Reader

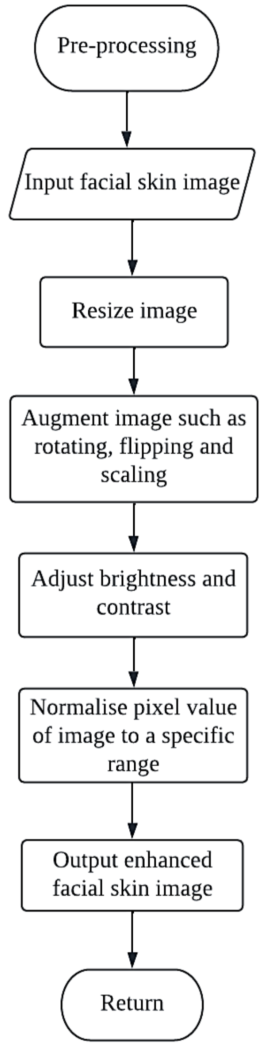


Figure 3. Flowchart of Pre-processing Phase

## IV. RESULTS AND DISCUSSIONS

This section presents the performance result of the YOLOv5 model, the result of skin type classification based on detected skin conditions, and the discussion.

### *A. Result of YOLOv5 Model*

The training of the YOLOv5 model was carried out with the training dataset and validation dataset. The YOLOv5 model was trained in nine skin conditions: acne, uneven skin, fine lines, blackheads, uneven skin tones, enlarged pores, shines, redness, and pigment. The total number for each class is shown in Figure 4. Various hyperparameters were tuned to train the YOLOv5 model with the most accurate skin conditions detection result. This is because hyperparameters can significantly affect the learning ability of the model, which is crucial for achieving excellent model performance. Metrics such as precision, recall, and mAP50 were employed to evaluate and compare the performance of the YOLOv5 model trained with different hyperparameters and frozen layers.

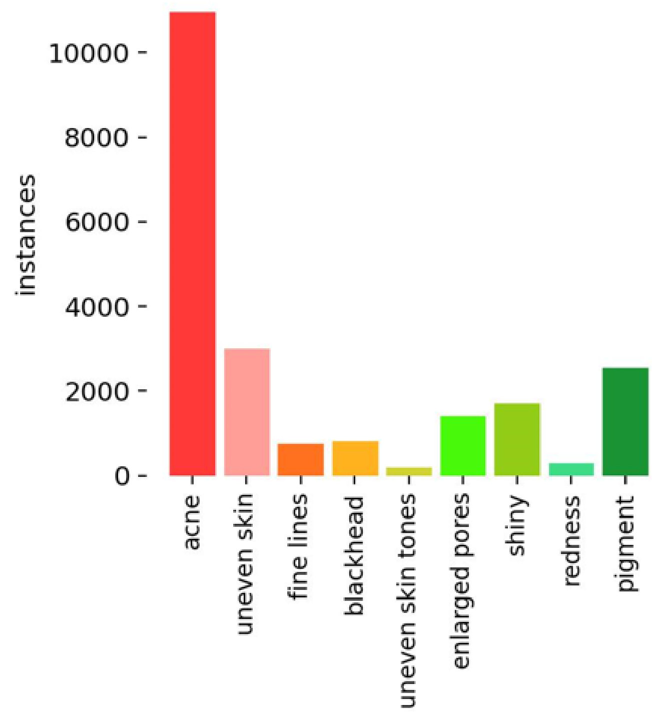


Figure 4. Total Number of Each Skin Condition

*i. Epoch*

The first hyperparameter tuned in the YOLOv5 model is the number of epochs. An epoch is a complete cycle of all the training datasets passing through the neural network. The number of epochs can be considered the total number of times the neural network is trained on the entire dataset. Adjusting the number of epochs can impact the accuracy and learning ability of the YOLOv5 model. In this paper, the YOLOv5 model was trained with 50, 80, 100, and 120 epochs. Table 1 shows the performance results obtained with a different epoch applied to all classes of skin conditions. The performance is evaluated in terms of precision, recall, and mAP50.

Table 1. Performance Results of Different Epoch for All Skin Conditions Classes

| Epoch | Precision | Recall | mAP50 |
|---|---|---|---|
| 50 | 0.481 | 0.378 | 0.383 |
| 80 | 0.492 | 0.387 | 0.377 |
| **100** | **0.540** | **0.446** | **0.424** |
| 120 | 0.471 | 0.404 | 0.386 |

As shown in the table above, the changes in precision, recall, and mAP50 were listed after training with 50, 80, 100, and 120 epochs. The 100 epochs produced the optimal result in all performance metrics compared to the 50, 80, and 120 epochs. The absolute precision, recall, and mAP50 found for 100 epochs were about 54%, 44.6%, and 42.4%, respectively.

*ii. Batch*

The second hyperparameter tuned in the YOLOv5 model is the batch size. Batch size is the number of images in the training dataset processed together in each iteration during training. Adjusting the batch size can impact the training process and convergence. In this paper, the YOLOv5 model was trained with 8, 16, and 24 batch sizes. The performance results obtained with different batch sizes applied to skin conditions are depicted in Table 2. The performance is evaluated in terms of precision, recall, and mAP50.

Table 2. Performance Results of Different Batch for All Skin Conditions Classes

| Batch Size | Precision | Recall | mAP50 |
|---|---|---|---|
| 8 | 0.467 | 0.401 | 0.383 |
| **16** | **0.540** | **0.446** | **0.424** |
| 24 | 0.487 | 0.405 | 0.381 |

The Table 2 shows how the performance metrics vary over multiple batch sizes. After training the YOLOv5 model several times, it was found that batch size 16 performed best in precision, recall, and mAP50, which were 54%, 44.6%, and 42.4%.

*iii. Performance Results of YOLOv5 in Skin Conditions Detection*

In this section, each performance result of YOLOv5 trained on various tuning hyperparameters is shown in the table below. It is clear from Table 3 that the YOLOv5 model with 100 epochs and a batch size of 16 is the best model for detecting skin conditions. It is said so because it achieves the highest accuracy and good precision, recall, and mAP50 performance. In short, this model achieved approximately 54% precision, 44.6% recall, and 42.4% mAP50 in detecting all categories of skin conditions. As shown in Figure 5, it can be seen from the trend that the precision, recall, and mAP50 of the YOLOv5 model with the epoch of 100 and batch size of 16 reached around 47% and above.

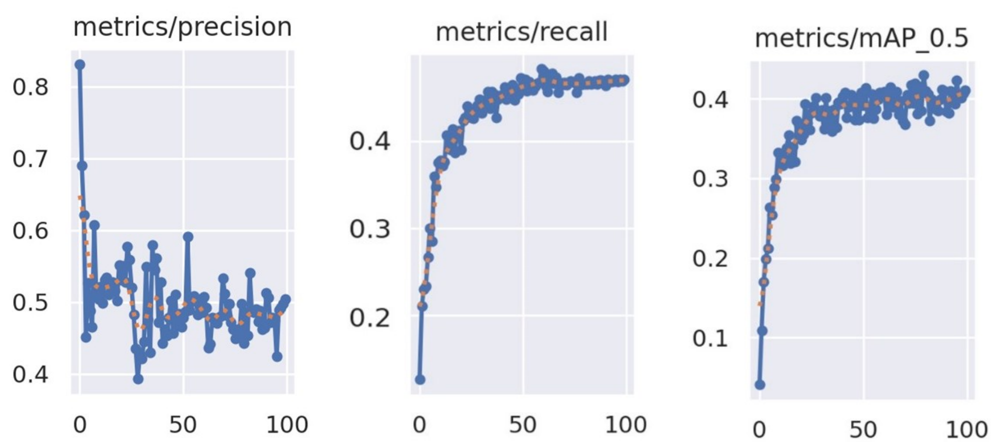


Figure 5. Performance Graphs of Precision, Recall, and mAP50 for 100 Epochs and 16 Batches

Table 3. Performance Results of YOLOv5 Model in Skin Conditions Detection

| Class | Fine-Tuned & Frozen Layer | Precision | Recall | mAP50 |
|---|---|---|---|---|
| **Acne** | 50 epochs | 0.540 | 0.430 | 0.450 |
| | 80 epochs | 0.516 | 0.424 | 0.433 |
| | **100 epochs** | **0.562** | **0.448** | **0.482** |
| | 120 epochs | 0.507 | 0.466 | 0.446 |
| | 8 batches | 0.513 | 0.464 | 0.453 |

| | | | | |
|---|---|---|---|---|
| | **16 batches** | **0.562** | **0.448** | **0.482** |
| | 24 batches | 0.502 | 0.462 | 0.447 |
| **Uneven Skin** | 50 epochs | 0.729 | 0.529 | 0.589 |
| | 80 epochs | 0.698 | 0.542 | 0.581 |
| | **100 epochs** | **0.748** | **0.691** | **0.721** |
| | 120 epochs | 0.683 | 0.574 | 0.604 |
| | 8 batches | 0.688 | 0.556 | 0.600 |
| | **16 batches** | **0.748** | **0.691** | **0.721** |
| | 24 batches | 0.675 | 0.555 | 0.589 |
| **Fine Lines** | 50 epochs | 0.615 | 0.500 | 0.529 |
| | 80 epochs | 0.602 | 0.509 | 0.532 |
| | **100 epochs** | **0.704** | **0.538** | **0.559** |
| | 120 epochs | 0.714 | 0.484 | 0.516 |
| | 8 batches | 0.640 | 0.505 | 0.517 |
| | **16 batches** | **0.704** | **0.538** | **0.559** |
| | 24 batches | 0.641 | 0.509 | 0.528 |
| **Blackhead** | 50 epochs | 0.441 | 0.273 | 0.246 |
| | 80 epochs | 0.322 | 0.234 | 0.233 |
| | **100 epochs** | **0.444** | **0.333** | **0.308** |
| | 120 epochs | 0.378 | 0.322 | 0.306 |
| | 8 batches | 0.357 | 0.301 | 0.251 |
| | **16 batches** | **0.444** | **0.333** | **0.308** |
| | 24 batches | 0.423 | 0.292 | 0.264 |
| **Uneven Skin Tones** | 50 epochs | 0 | 0 | 0.044 |
| | 80 epochs | 0.387 | 0.0147 | 0.0647 |
| | **100 epochs** | **0.388** | **0.060** | **0.0768** |
| | 120 epochs | 0.137 | 0.0517 | 0.0553 |
| | 8 batches | 0.107 | 0.0294 | 0.0641 |
| | **16 batches** | **0.388** | **0.060** | **0.0768** |
| | 24 batches | 0.274 | 0.0557 | 0.0617 |
| **Enlarged Pores** | 50 epochs | 0.425 | 0.350 | 0.313 |
| | 80 epochs | 0.427 | 0.365 | 0.299 |
| | **100 epochs** | **0.435** | **0.415** | **0.323** |
| | 120 epochs | 0.352 | 0.386 | 0.284 |
| | 8 batches | 0.402 | 0.412 | 0.316 |
| | **16 batches** | **0.435** | **0.415** | **0.323** |
| | 24 batches | 0.415 | 0.409 | 0.322 |

| | | | | |
|---|---|---|---|---|
| **Shiny** | 50 epochs | 0.473 | 0.433 | 0.379 |
| | 80 epochs | 0.404 | 0.449 | 0.356 |
| | **100 epochs** | **0.476** | **0.468** | **0.388** |
| | 120 epochs | 0.453 | 0.466 | 0.391 |
| | 8 batches | 0.442 | 0.424 | 0.388 |
| | **16 batches** | **0.476** | **0.468** | **0.388** |
| | 24 batches | 0.425 | 0.424 | 0.365 |
| **Redness** | 50 epochs | 0.435 | 0.319 | 0.290 |
| | 80 epochs | 0.431 | 0.348 | 0.287 |
| | **100 epochs** | **0.430** | **0.392** | **0.291** |
| | 120 epochs | 0.356 | 0.217 | 0.207 |
| | 8 batches | 0.429 | 0.283 | 0.244 |
| | **16 batches** | **0.430** | **0.392** | **0.291** |
| | 24 batches | 0.408 | 0.304 | 0.215 |
| **Pigment** | 50 epochs | 0.673 | 0.570 | 0.609 |
| | 80 epochs | 0.636 | 0.595 | 0.602 |
| | **100 epochs** | **0.674** | **0.672** | **0.664** |
| | 120 epochs | 0.656 | 0.670 | 0.661 |
| | 8 batches | 0.622 | 0.631 | 0.616 |
| | **16 batches** | **0.674** | **0.672** | **0.664** |
| | 24 batches | 0.622 | 0.634 | 0.634 |

### *B. Result of Skin Type Classification*

The system's function and skin type classification are achieved by detecting skin conditions. Each of the skin types has its characteristics. Therefore, the user's skin type can be known based on the skin conditions detected by the YOLOv5 model. For example, dry skin is less likely to have acne and oiliness, uneven skin, blackhead, uneven skin tones, and enlarged pores. In contrast, oily skin will have a greater chance of developing acne blemishes, uneven skin, blackheads, uneven skin tones, and enlarged pores. The t-zone of the oily face is also shiny. Moreover, normal skin does not suffer from skin problems, while sensitive skin is more prone to skin irritation and redness.

This system has specified various conditions to serve as a baseline for each skin type. A skin is considered dry if the acne is three or less, whereas if acne is four or more, it is classified as oily skin. Besides that, if no acne is detected, the skin is deemed normal, and the skin is termed sensitive if there is facial redness. As shown in Figure 6, 7, 8, and 9, some facial skin images were tested to prove the system can provide accurate skin type classification results according to detected skin conditions.

As shown in Figure 6, most skin problems are detected and correctly identified. It classified the user's facial skin as dry. This output is triggered because the skin conditions detected have only a single acne, two oiliness, and without redness, which does not meet the characteristics of oily and sensitive skin. It is also unsuitable for normal skin because there is acne. Thus, it is undeniable that the system can accurately recognize if the user's facial skin is dry.

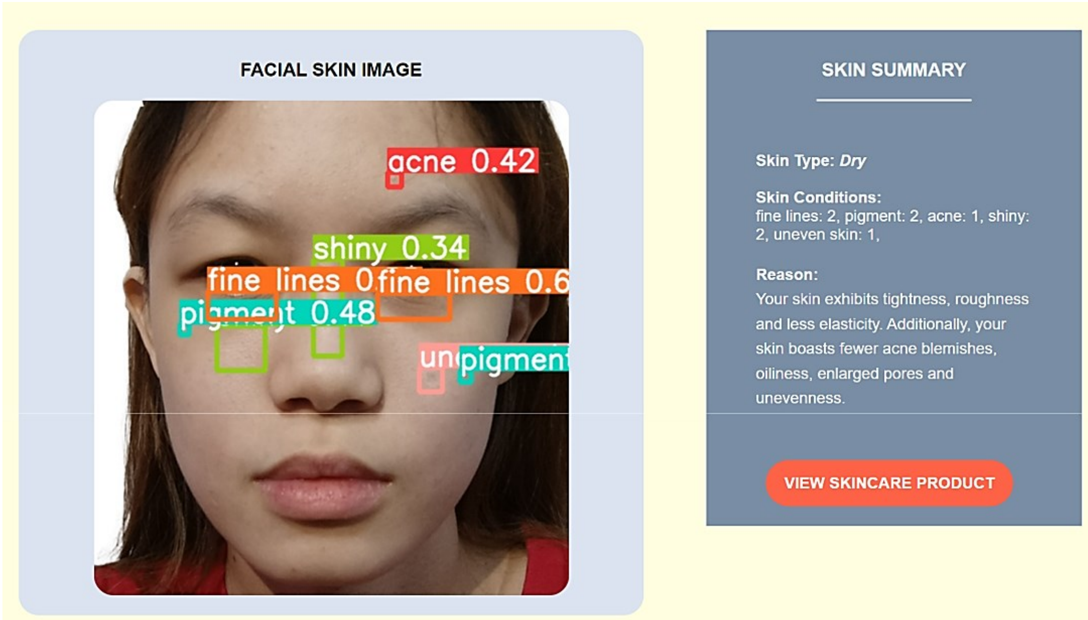


Figure 6. Dry Skin

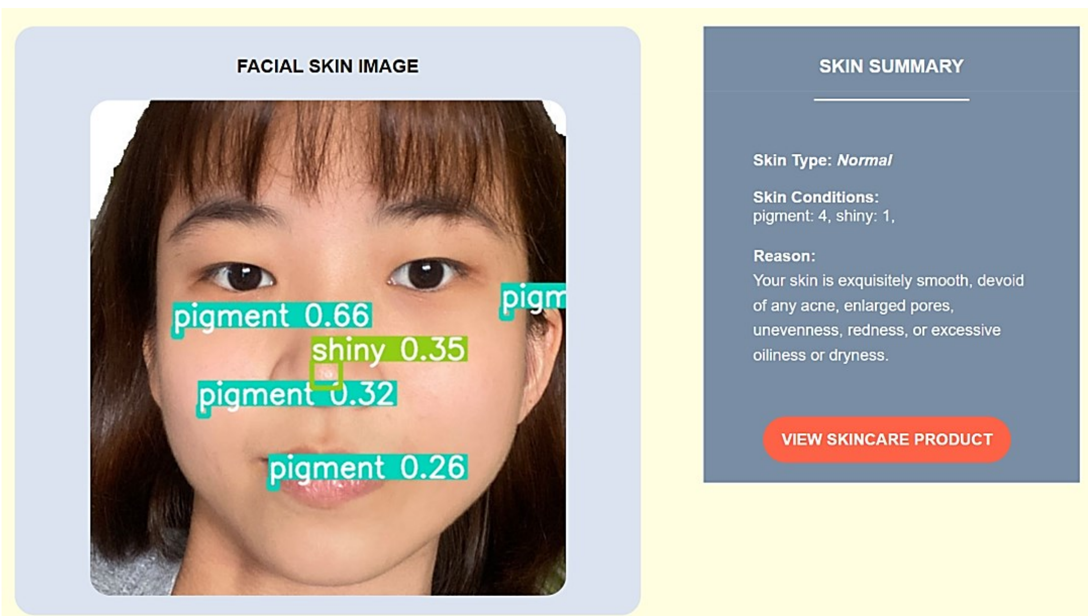


Figure 7. Normal Skin

As shown in Figure 7, most skin problems are detected and correctly identified. It classified the user's facial skin as normal. This output is triggered because the identified skin conditions do not have a skin issue incompatible with dry, oily, and sensitive skin features. The system can identify whether the user's facial skin condition is normal.

As shown in Figure 8, most skin problems are detected and correctly identified. It classified the user's facial skin as oily. This output is triggered because the detected skin conditions have more than four acnes, and the skin is shiny and uneven, inconsistent with dry and normal skin characteristics. Since no redness was observed on the skin, it was not classified as sensitive. Undeniably, the system can properly determine whether the user's facial skin is oily.

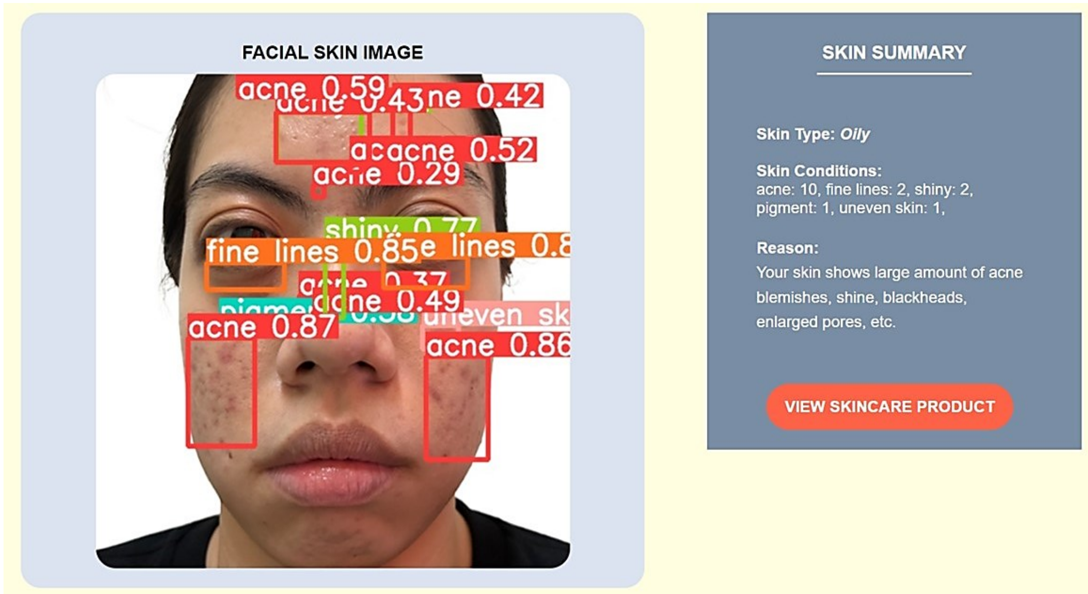


Figure 8. Oily Skin

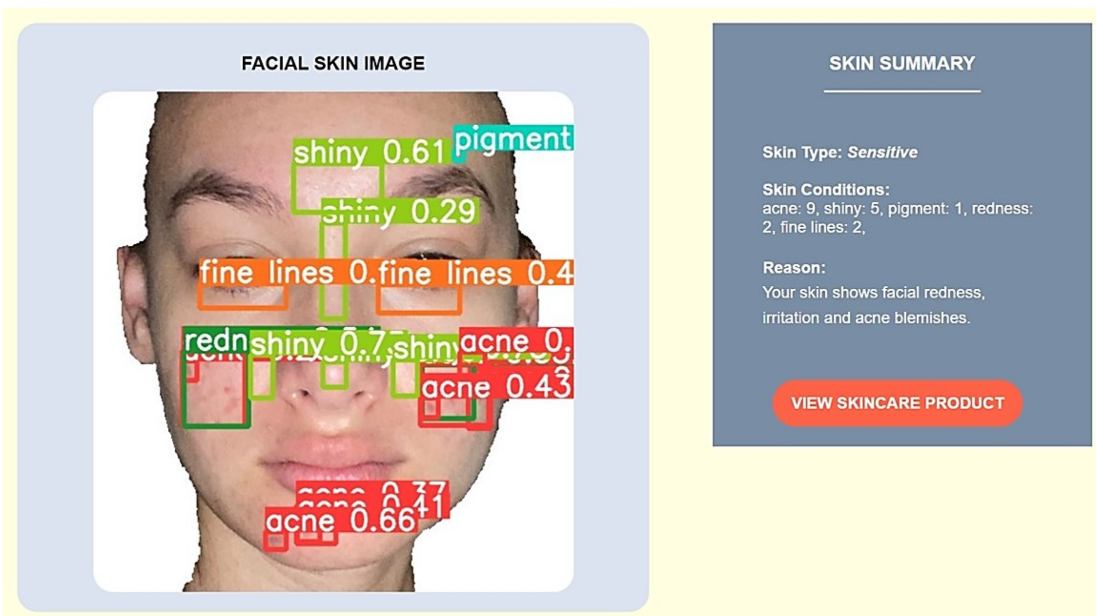


Figure 9. Sensitive Skin

As shown in the Figure 9, most skin problems are detected and correctly identified. It classified the user's facial skin as sensitive. This output is triggered because the skin conditions detected have more than one redness, incompatible with dry and normal skin features. Although many acne and oiliness are detected, it is also not ideal for oily skin because it possesses facial redness. The system can accurately recognize if the user's facial skin is sensitive.

*C. Discussions*

The results of the YOLOv5 model trained on different hyperparameters indicate that higher epochs and batch sizes are more likely to yield good performance in detecting skin conditions. Although adding more epochs may lengthen training time, the detection performance can be improved because the neural network will have more chances to learn the underlying patterns in the data without overfitting. The higher the epochs, the higher the opportunity for the model to learn and refine its biases and weights. Monitoring the performance results of the YOLOv5 model shows that the optimal number of epochs is 100 in detecting skin conditions. The 50 and 80 epochs are too few, leading to underfitting, in which the neural network cannot adequately grasp the data's complexity. Also, the 120 epoch is too large, which results in overfitting because the neural network gets highly specialized to the training data and underperforms on data that has not been seen before.

It may affect the training process's convergence and stability regarding the batch size. Larger batch sizes can result in more stable and smoother gradients and quicker convergence during training. Also, more images processed concurrently help speed up gradient calculations and weight updates. On the other hand, smaller batch sizes increase stochasticity and can improve the model's generalization by exploring a broader range of images throughout each iteration. In this case, the model must be trained multiple times to know the most suitable batch size. After experiments with different batch sizes, the performance results of the YOLOv5 model indicated that the most appropriate batch size is 16. Moreover, training speed can be accelerated by using a bigger batch size. However, it will require more GPU memory, which may be a constraint depending on the resources available on the computer.

## V. CONCLUSION

A skin type classification system has been introduced in this paper. The facial skin analysis was mainly developed based on Malaysian skin using YOLOv5 technology. YOLOv5 can detect users' skin conditions and classify their skin type, then transform the prediction result to the product recommendation stage to determine the skincare product most suitable for the user's skin type.

Malaysians' facial skin images were used as the datasets for the YOLOv5 model. The experiments showed that the precision, recall, and mAP50 of YOLOv5 in skin conditions detection reached approximately 54%, 44.6%, and 42.4%. The main factor that hinders the performance of this system is the limitation and insufficient dataset of Malaysian facial skin. With this promising preliminary result, we postulated that the performance would be much improved by increasing the dataset. Thus, collecting a dataset of Malaysian skin types would also be a promising research gap to work into. On top of this, we also noticed that a more extensive epoch and batch size performed better than a smaller size for the YOLOv5 model to detect skin conditions.

## ACKNOWLEDGEMENT

A version of this paper was presented at the third International Conference on Computer, Information Technology and Intelligent Computing, CITIC 2023, held in Malaysia on 26th-28th July 2023.

The authors received no funding from any party for the research and publication of this article.

## AUTHOR CONTRIBUTIONS

Ying Huey Gan: Data Curation, Formal Analysis, Investigation, Visualization, Writing – Original Draft Preparation;
Shih Yin Ooi: Project Administration, Resources, Supervision, Writing – Review & Editing;
Ying Han Pang: Supervision, Writing – Review & Editing;
Yi Hong Tay: Methodology, Validation, Visualization;
Quan Fong Yeo: Methodology, Validation, Investigation

## CONFLICT OF INTERESTS

No conflict of interests were disclosed.

## REFERENCES

[1] O.K. Pal (2021). Skin Disease Classification: A Comparative Analysis of K-Nearest Neighbors (KNN) and Random Forest Algorithm. Proceedings of International Conference on Electronics, Communications and Information Technology, ICECIT 2021. https://doi.org/10.1109/ICECIT54077.2021.9641120

[2] T. Parashar, K. Joshi, R.N. Ravikumar, D. Verma, N. Kumar, & K.S. Krishna (2022). Skin Disease Detection using Deep Learning. Proceedings of the 2022 11th International Conference on System Modeling and Advancement in Research Trends, SMART 2022, 1380–1384. https://doi.org/10.1109/SMART55829.2022.10047465

[3] K.V. Swamy, & B. Divya (2021). Skin Disease Classification using Machine Learning Algorithms. Proceedings of the 2021 2nd International Conference on Communication, Computing and Industry 4.0, C2I4 2021. https://doi.org/10.1109/C2I454156.2021.9689338

[4] V. Vasudha Rani, G. Vasavi, & B. Maram (2022). Skin Disease Classification Using Machine Learning and Data Mining Algorithms. Proceedings - 2022 IEEE 2nd International Symposium on Sustainable Energy, Signal Processing and Cyber Security, ISSSC 2022. https://doi.org/10.1109/ISSSC56467.2022.10051620

[5] H.T. Chan, Y.W. Liao, S.Y. Jhong, S.C. Chien, K.L. Hua, & Y.Y. Chen (2023). A Skin Type Classification Method Using Mobile Device-Based Deep Learning Model. 2023 9th International Conference on Applied System Innovation, ICASI 2023, 199–201. https://doi.org/10.1109/ICASI57738.2023.10179572

[6] D. Hecht, T. Pfahler, I. Ullmann, T. Altstidl, N. Amer, Y. Jin, B. Eskofier & M. Vossiek (2022). In Vivo Skin-Type Classification Using Millimeter-Wave Near-Field Probe Spectroscopy. 2022 52nd European Microwave Conference, EuMC 2022, 404–407. https://doi.org/10.23919/EUMC54642.2022.9924376

[7] J.C. Dunn (1973). A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. 3(3), 32–57. https://doi.org/10.1080/01969727308546046

[8] J.C. Bezdek (1981). Pattern Recognition with Fuzzy Objective Function Algorithms. https://doi.org/10.1007/978-1-4757-0450-1

[9] J.-d. Lee, Y.-c. Chen, L.-c. Liu, & C.-h. Huang (2007). A Facial-skin Condition Classification System Using Texture and Fuzzy C-means Technologies. https://doi.org/10.1109/FSKD.2007.22

[10] L.A. Zadeh (1965). Fuzzy Sets. https://doi.org/10.1016/S0019-9958(65)90241-X

[11] Y.P. Huang, Y.Z. Li, & C.C. Lin (2013). Smart Phone-Based Fuzzy Modeling to Examine Facial Skin Quality. ICSSE 2013 - IEEE International Conference on System Science and Engineering, Proceedings, 69–74. https://doi.org/10.1109/ICSSE.2013.6614635

[12] C. Cortes, & V. Vapnik (1995). Support-Vector Networks Editor. In Machine Learning (Vol. 20). https://dx.doi.org/10.1007/BF00994018

[13] T.Y. Lin, H.T. Chan, C.H. Hsia, & C.F. Lai (2022). Facial Skincare Products' Recommendation with Computer Vision Technologies. Electronics (Switzerland), 11(1). https://doi.org/10.3390/electronics11010143

[14] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner (1998). Gradient-Based Learning Applied to Document Recognition. https://doi.org/10.1109/5.726791

[15] A. Kothari, D. Shah, T. Soni, & S. Dhage (2021). Cosmetic Skin Type Classification Using CNN With Product Recommendation. 2021 12th International Conference on Computing Communication and Networking Technologies, ICCCNT 2021. https://doi.org/10.1109/ICCCNT51525.2021.9580174

[16] K. He, X. Zhang, S. Ren, & J. Sun (2015). Deep Residual Learning for Image Recognition. https://doi.org/10.48550/arXiv.1512.03385

[17] S. Sharma (2020). DermaDetect: A Novel Computer Vision Model for An Accurate Diagnosis of Skin Conditions and Rashes. Proceedings - 2020 IEEE 7th International Conference on Data Science and Advanced Analytics, DSAA 2020, 743–744. https://doi.org/10.1109/DSAA49011.2020.00094

[18] G. Jocher, A. Chaurasia, A. Stoken, J. Borovec, NanoCode012, Y. Kwon, K. Michael, TaoXie, J. Fang, imyhxy, Lorna, Y. Zeng, C. Wong, V. Abhiram, D. Montes, Z. Wang, C. Fati, J. Nadar, Laughing, UnglvKitDe, V. Sonck, tkianai, yxNONG, P. Skalski, A. Hogan, D. Nair, M. Strobel, M. Jain. (2022). ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation. https://doi.org/10.5281/zenodo.7347926

[19] J. Chhatlani, T. Mahajan, R. Rijhwani, A. Bansode, & G. Bhatia (2022). DermaGenics - Early Detection of Melanoma using YOLOv5 Deep Convolutional Neural Networks. 2022 IEEE Delhi Section Conference, DELCON 2022. https://doi.org/10.1109/DELCON54057.2022.9753227

[20] J.J. Ng, K.O.M.Goh, & C.Tee (2023). Traffic Impact Assessment System using Yolov5 and ByteTrack. Journal of Informatics and Web Engineering. 2(2), 168-188. https://doi.org/10.33093/jiwe.2023.2.2.13

[21] J.-R. Lee, K.-W, Ng, & Y.-J, Yoong (2023). Face and Facial Expressions Recognition System for Blind People Using ResNet50 Architecture and CNN. Journal of Informatics and Web Engineering. 2(2), 284-298. https://doi.org/10.33093/jiwe.2023.2.2.20

[22] Q. F. Yeo, S. Y. Ooi, Y. H. Pang, and Y. H. Gan, "Facial Skin Type Analysis Using Few-shot Learning with Prototypical Networks", International Journal on Advanced Science, Engineering and Information Technology, vol. 13, no. 6, pp. 2249–2266, Dec. 2023.