



# Multi-agent learning neural network and Bayesian model for real-time IoT skin detectors: a new evaluation and benchmarking methodology

A. A. Zaidan<sup>1</sup> · B. B. Zaidan<sup>1</sup> · M. A. Alsalem<sup>2</sup> · O. S. Albahri<sup>1</sup> · A. S. Albahri<sup>3</sup> · M. Y. Qahtan<sup>1</sup>

Received: 1 June 2018 / Accepted: 22 June 2019 / Published online: 2 July 2019  
© Springer-Verlag London Ltd., part of Springer Nature 2019

## Abstract

This study aimed to develop a new methodology for evaluating and benchmarking a multi-agent learning neural network and Bayesian model for real-time skin detectors based on Internet of things (IoT) by using multi-criteria decision-making (MCDM). The novelty of this work is in the use of an evaluation matrix for the performance evaluation of real-time skin detectors that are based on IoT. Nevertheless, an issue with the performance evaluation of real-time skin detector approaches is the determination of sensible criteria for performance metrics and the trade-off amongst them on the basis of different colour spaces. An experiment was conducted on the basis of three phases. In the first phase, a real-time camera based on cloud IoT was used to gather different caption images. The second phase could be divided into two stages. In the first stage, a skin detection approach was developed by applying multi-agent learning based on different colour spaces. This stage aimed to create a decision matrix of various colour spaces and three groups of criteria (i.e. reliability, time complexity and error rate within a dataset) for testing and evaluating the developed skin detection approaches. In the second stage, Pearson rules were utilised to calculate the correlation between the criteria in order to make sure, either needs to use all of the criteria in decision matrix and the criteria facts that affect the behaviour of each criterion, in order to make sure that use all the criteria in evaluation as multidimensional measurements or not. In the third phase, the MCDM method was used by integrating between a technique in order of preference by similarity to the ideal solution and multi-layer analytic hierarchy process to benchmark numerous real-time IoT skin detection approaches based on the performed decision matrix from the second phase. Three groups of findings were obtained. Firstly, (1) statistically significant differences were found between the criteria that emphasise the need to use all of the criteria in evaluation. (2) The behaviour of the criteria in all scenarios was affected by the distribution of threshold values for each criterion based on the different colour spaces used. Therefore, the differences in the behaviour of criteria that highlight the use of the criteria in evaluation were included as multidimensional measurements. Secondly, an overall comparison of external and internal aggregation values in selecting the best colour space, namely the normalised RGB at the sixth threshold, was discussed. Thirdly, (1) the YIQ colour space had the lowest value and was the worst case, whereas the normalised RGB had the highest value and was the most recommended of all spaces. (2) The lowest threshold was obtained at 0.5, whereas the best value was 0.9.

**Keywords** Multi-criteria decision-making techniques · Skin detector within IoT · Evaluation and benchmarking multi-criteria analysis

✉ A. A. Zaidan  
aws.alaa@gmail.com; aws.alaa@fskik.upsi.edu.my

B. B. Zaidan  
bilalbahaa@fskik.upsi.edu.my

M. A. Alsalem  
mohammed.asum@gmail.com

O. S. Albahri  
osamahsh89@gmail.com

A. S. Albahri  
ahmed.bahri1978@gmail.com

M. Y. Qahtan  
yahoopd@gmail.com

Extended author information available on the last page of the article

## 1 Introduction

The Internet of things (IoT) is a rapidly developing innovation that substantially affects the way humans live [1]. IoT is ‘a global infrastructure for the information society, enabling advanced services by interconnecting (physical and virtual) things based on existing and evolving interoperable information and communication technologies’ [2–6]. IoT provides capabilities that authorise the communication of devices with one another and the automatic real-time exchange of information [7]. IoT is widely used in several fields for different applications [8–11]. Image processing is an IoT application that is reaching a distinct level and becoming a major part of people’s lives. IoT applications have created many trends in the image processing field [12]. Consequently, IoT and image processing have been independently used in different applications [5]. IoT provides the ease of monitoring and gaining image data through the Internet [13]. The need for high-quality image data is increasing to the extent that real-time skin detection is required [9, 14]. Real-time skin detection within IoT is implemented on the basis of artificial intelligence (AI) algorithms, which can detect human skin by using cameras in real time [15]. Thus, real-time skin detection within IoT based on AI algorithms involves functional and efficient mechanisms during segmentation, and many studies have applied various adaptability strategies for soft computing techniques [16, 17]. Real-time skin detection within IoT is important in various applications, such as gesture analysis, facial analysis, human–machine interface, image content filter, video surveillance, annotation and colour-balancing applications [16, 18]. With the rapid development of skin detection approaches in different applications and their recent integration with IoT, finding a reliable, effective and comprehensive evaluation and benchmarking methodologies has become essential. Evaluating the performance of an application based on IoT services is a challenge because many considerations and factors are included. The acquired data in IoT may not meet the requirements of systems because of the following factors. Firstly, data in practice are commonly noisy because of the environmental noise and sensing devices’ deviation and limitation of sensing accuracy. Deviation and sensing accuracy may differ from device to device (D2D). Secondly, data may be corrupted by malicious data. Thirdly, data can be incomplete; as such, the numerical limit of sensor devices and constrained sensing cost must be considered. Fourthly, even accurate and complete data can be outdated for the demand [14, 19]. Many studies have dealt with different criteria commonly used in IoT-based skin detection. For example, reliability and time complexity have been utilised [19, 20], whereas reliability criteria have been considered only to evaluate

results in a system-based IoT [21, 22]. According to [23], the design of any skin detection approach should adapt to the criteria (reliability/time complexity/dataset). Early studies emphasised the problems of skin detection evaluation and benchmarking and mentioned three general requirements, namely (1) reliability (i.e. obtained skin detection rate and false positives [FPs]), (2) time complexity and (3) datasets (i.e. obtained equal error rate from a histogram classifier). With the aid of contrast, the time-consuming observation of histogram model performance has been observed. In another study [24], the dataset criterion is highlighted by comparing two algorithms despite the significance of the remaining criteria. For skin segmentation images of nonskin and skin pixels, the dataset is presented by training and testing. However, the classifier that creates output images is compared pixelwise with the ground truth of skin segmentation. Consistent with [25], our study reports a skin detection algorithm examined with images from independent databases. In general, image size must be considered for the evaluation of time complexity. This research shows that increasing the image accuracy results in the enhancement of training data, thereby increasing the time complexity of the experiment. Reliability is a prerequisite for the evaluation of skin detection [26]. Studies have emphasised that the reliability criterion primarily depends on accuracy, precision and recall of image colour despite the significance of the remaining criteria. Nonetheless, further high-quality assessment of skin detection is needed. Two parameters, namely reliability and dataset, have been reported [27]. A relationship between reliability and dataset criteria has also been described. In particular, the computation of reliability depends on a large dataset with a manually described ground truth by using an ROC curve based on the testing methodology. By contrast, time complexity and reliability have been studied, and their effects on each other have been emphasised [28]. Some studies have argued that their relationship depends on an adopted algorithm’s complex background, whereas other studies have highlighted that obtaining high accuracy and excellent quality throughout the duration of skin detection is based on processor computation features. A previous study [29] proposed three individually evaluated criteria for skin detection and discussed their effects. In another study [24], the basic criteria for reliability computation (i.e. reliability group) include training, testing (i.e. dataset group) and time complexity which are used to evaluate and benchmark skin detection approaches. Consequently, the major challenge in creating a skin detection approach is the conflict between skin detection evaluation criteria. Therefore, these requirements must be considered by evaluation and benchmarking. In other studies [23–26], all the proposed criteria have been adopted regardless of the trade-off amongst them. These

studies have also evaluated the reliability criterion for a given time complexity dependent on diverse datasets and shown that the percentage of reliability varies in terms of the adopted algorithm. Therefore, the reliability criterion no longer has a constant level. The variety in time complexity is stated amongst algorithms that are dependent on the CPU time [30]. For any image within the scope of the present research, processing time is an important factor in evaluation process. Therefore, calculations must have the highest percentage of reliability when it is compared with the lowest time complexity of the output image. A dataset can be classified into two, namely training and validation data classes, to find a minimum detection error [31]. Thus, these studies confirmed the evaluation process for each criterion based on independent guidelines.

Previous studies also suggested the reliability criterion but did not refer to a specific level during the comparison between different criteria. Therefore, conflicting criteria or trade-off problems amongst reliability, time complexity and error rate within a dataset of a skin detector approach are clearly reported in prior studies [27]. Notably, problems on evaluation in skin detection are defined as multi-criterion problems with conflicting criteria. However, the following must be ensured. (1) All the available criteria used are in evaluation matters, and (2) all the criteria in evaluation are used as multidimensional measurements or not. Thus, the performance of criteria is highlighted by finding the relationship between criteria, calculating the correlation coefficient and identifying the behaviour of different criteria based on different thresholds with various colour spaces, and these processes are necessary to design a new methodology for the testing, evaluating and benchmarking of real-time IoT skin detection approaches. Moreover, the new methodology should be flexible and capable of handling the conflicting criterion problems after the current criterion needs are maintained [32–34]. Therefore, using structured and explicit approaches in decisions that involve multiple attributes can improve the quality of decision-making [35–37], and a set of techniques, which is classified under the collective heading multiple criteria decision analysis (MCDA), is useful for this purpose. MCDA is a sub-discipline of operational research and explicitly considers multiple criteria in decision-making conditions that occur in various actual situations in different domains [2]. Several useful techniques can be utilised to address multi-attribute decision-making or multi-criterion decision-making (MADM/MCDM) problems in the real world [38–42]. These methods not only help decision-makers (DMs) organise problems but also analyse, rank and score alternatives [2, 3, 43, 44]. Accordingly, the scoring of a suitable alternative should be performed. MADM/MCDM methods can solve benchmarking problems for real-time IoT skin detectors. In any MADM/MCDM benchmarking,

fundamental terms should be defined in a decision matrix form that includes  $m$  alternatives and  $n$  criteria. A crossover between all criteria and alternatives is represented by  $x_{ij}$ . Ultimately, we obtain the  $(x_{ij})_{(m*n)}$  matrix, which is expressed as follows:

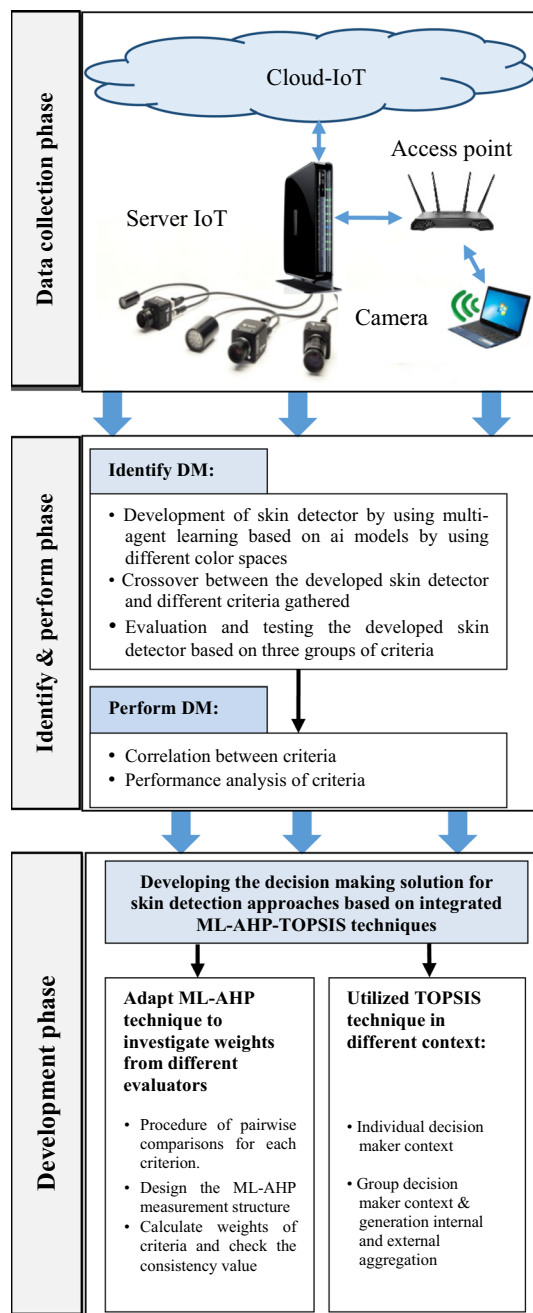
$$D = \begin{matrix} & C_1 & C_2 & \dots & C_n \\ & W_1 & W_2 & \dots & W_n \\ A_1 & \left[ \begin{array}{cccc} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{array} \right] \\ A_2 & & & & \\ \vdots & & & & \\ A_m & & & & \end{matrix},$$

where  $A_1, A_2, \dots, A_m$  are the possible alternatives that decision-makers must score (i.e. approaches of skin detection);  $C_1, C_2, \dots, C_n$  are the criteria against which the overall performance of all alternatives is measured (i.e. reliability, time complexity and error rate within the data set);  $W_1, W_2, \dots, W_n$  are the weight of the criteria;  $x_{ij}$  is the alternative rating and  $A_i$  with respect to criterion  $C_j$ ; and  $W_j$  is the weight of criterion  $C_j$ . Certain techniques, such as normalisation, use of maximisation indicators and addition of weights, need to be finished to rank each alternative [45, 46].

In this study, we present a new methodology to aid the decision-making process for evaluating and benchmarking IoT real-time skin detectors based on a multi-agent learning neural network and a Bayesian model through multi-criterion analysis. The novelty of this study is in the utilisation of an evaluation matrix for the performance evaluation of real-time skin detectors that are based on IoT. The remaining sections of this paper are organised as follows. In Sect. 2, we describe the methodology. In Sect. 3, we present the results and discussion. In Sect. 4, we discuss the validation of the results of the proposed methodology. In Sect. 5, we draw our conclusions.

## 2 Methodology

The proposed methodology is conducted on the basis of three phases (Fig. 1). In the first phase, data are collected from a real-time camera within cloud IoT to gather different images. In the second phase, the process adaptation of the best previous case of the skin detector is identified using multi-agent learning, depending on different colour spaces, to create a dataset of different colour space samples for benchmarking and to perform a crossover between the adopted skin detector and the multi-evaluation group's criteria (i.e. reliability, time complexity and error rate within dataset) of the groups gathered for decision-making. The evaluation of the adopted skin detector and the testing are dependent on three groups of criteria. This phase is also conducted to analyse the performance between different



**Fig. 1** Methodology of the experimental phases for evaluating and benchmarking the real-time IoT skin detector

criteria based on the data of the criteria. This phase mainly aims to determine the importance of finding the relationship between the different data of the criteria in this phase. This operation is implemented to verify the existing statistical differences between them; otherwise, only one is used. The Pearson formula is adapted to calculate the correlation coefficient between various criteria, to investigate the relationship between the criteria and determine their degree of correlation. A performance analysis is conducted to evaluate and compare the criteria and identify the factors

that affect their behaviour. In the third phase, a new MCDM approach is developed and subsequently used with the integrated TOPSIS (technique in order of preference by similarity to the ideal solution) and ML-AHP (multi-layer-analytic hierarchy process) as a basis for benchmarking several skin detection approaches by using the performed decision matrix outcome from the second phase.

## 2.1 Data collection phase

This phase includes a range of devices and sensors used to monitor and capture images within the IoT system. This system monitors and captures images across a range of cameras when motion is detected. Data are collected in the form of images by cameras scattered in specific locations. These cameras are connected to the Internet and send the captured images to the main centre, which is represented by a central server [47–49].

The server collects the data received from the surveillance camera to be processed and configured and then sends the data to terminals via access points deployed in certain locations [50–52]. These routers transfer data through Wi-Fi networks [2, 52, 53]. Computers receive the data to be stored and then perform the required processing. Subsequently, a computer vision module is applied to the captured images to reveal the skin, and skin detector approaches are evaluated and benchmarked in the succeeding phase.

## 2.2 Identification and performance phase

This phase includes two main stages, namely identification of a decision matrix and performance of the decision matrix. These stages are discussed in detail in the following subsections.

### 2.2.1 Identification of the decision matrix

Ascertaining skin detector approaches is an important stage in the creation of the decision matrix and comprises three main steps: developing the skin detector by using different colour spaces, conducting a crossover between different criteria with developed skin detector engines and evaluating the developed skin detector and testing it against the criteria of the three groups. This stage establishes the decision matrix from a practical aspect, which is explored in detail in the following subsections.

**2.2.1.1 Development of skin detector by using multi-agent learning dependent on different colour spaces** This section highlights a case study adopted in the current work. The case study is developed on the basis of the selected colour spaces. This step is important in completing the

identification and performance phase. The development of multi-agent learning technique for the skin detector is discussed in detail.

**2.2.1.1.1 Multi-agent learning technique** In a new method adopted in a previous study [54], multi-agent learning is used to resolve the three key issues in skin detection. This study solves a skin-like problem by utilising the most appropriate method in parametric skin modelling (NN model) with a segment adjacent-nested (SAN) technique. By contrast, the present work involves the most appropriate nonparametric method in skin modelling (Bayesian model) with a grouping histogram (GH) technique to resolve lighting condition problems. Subsequently, the group combines the two models to solve the reflection problem with water and glass. The NN method resolves skin problems. Conversely, the Bayesian model fails to resolve the lighting condition entirely. Therefore, this experiment is conducted using different colour spaces with the Bayesian model to improve results in a particular model.

**2.2.1.1.2 Adapted colour space** Fourteen samples of different colour spaces obtained from previous studies are used in the present study [55]. The procedure applied to the colour groups provides a solution to the lighting condition problem, which depends on the removal of the illumination component from these samples by using the Bayesian model of the proposed multi-agent learning technique. Colour spaces are widely used in studies on skin detection because they address many problems in this field [32, 54, 56–64]. In our study, each colour space is built depending on the separation of the illumination element from the chroma element, and this process is a key step in the development phase. Therefore, the luminance element is ignored, whereas the chroma element is retained because it is necessary to determine the skin colour during skin detection. Hence, skin detection is conducted because of the elimination of the luminance component, which is an important aspect in the size reduction in skin clusters in a colour space [55, 65].

Different colour spaces are discussed in detail in the following section.

- *Normalised RGB*

Different colour spaces can be altered to represent RGB easily. The components of RGB can represent colour and luminance and are subsequently used to represent skin colour in their chromatic colour space. Luminance can be isolated from the colour space through normalisation. Chromatic colours, which are also known as ‘pure’ colours, in the absence of luminance are defined as follows:

$$R = R/(R + G + B), \quad G = G/(R + G + B).$$

The process removes B, which represents a component of luminance [66–68].

- *YCbCr*

YCbCr is ‘an encoded nonlinear RGB signal generally used by European television studios and utilised for image compression patterns’. This colour space is a clear option for skin detection because it efficiently separates luminance and easily converts from RGB and vice versa.

$$Y = 0.299R + 0.587G + 0.114B, \quad Cb = B - Y, \\ Cr = R - Y$$

Y represents the excluded luminance component whilst using consistency ratio (Cr) and Cb only [69, 70].

- *YCgCr*

YCgCr is ‘a colour space derived from YCbCr and has a Y channel that provides the luminous component, that is, light intensity; meanwhile, Cg and Cr channels represent the green and red differences in chromaticity components, respectively’. This colour space is used for digital video encoding. Y is averted, and only the chrominance fraction is adopted in the proposed integrated approach [71, 72].

The YCgCr colour space is generated through the transformation of the RGB values by using the following equations:

$$Y = 16 + 65.481R + 128.553G + 24.966B, \\ Cg = 128 - 81.085R + 112G - 30.915B, \\ Cr = 128 + 112R - 93.768G - 18.214B.$$

- *YCgCb*

YCgCb is ‘another colour space derived from YCbCr, and the RGB image determines the fitting skin region for each Y as the luminance component and the two chrominances as Cg and Cb’. The luminescent factor is excluded, whereas the chrominance element is retained. If the colour components of a pixel are within the boundaries of a fitting skin region, then this pixel is classified as a skin pixel. The Cg–Cb colour space for skin tone detection is represented by a circular model [73].

Therefore, the circular model for the skin tone in the transformed Cg–Cb space is portrayed as follows:

$$\frac{(x - Cg)^2 + (y - Cb)^2}{12.25^2}, \quad \left( \begin{array}{l} x = 12.25 \cos + 107 \\ y = 12.25 \sin - 110 \end{array} \right).$$

- *YUV*

In this colour space, Y is ‘the luminance component, and UV is represented as the chrominance component’. In

YUV, a colour space removes the luminance-related component Y to enhance the performance of the skin detection process. As such, the definition of the luminance component in this colour space is a good step towards obtaining invariance to luminance. Human vision is critical to the luminance and chrominance factors in images. The colour space confirms this sensitivity by increasing the bandwidth of the luminance to be close to human perception. YUV is derived from the original RGB source. Thus, this colour space can be converted to RGB through linear transformations [72, 74].

The YUV colour space is generated by the transformation of the RGB values by using the following equations:

$$Y = +0.299R + 0.587G + 0.114B,$$

$$U = -0.14713R - 0.28886G + 0.436B,$$

$$V = +0.615R - 0.51499G - 0.10001B.$$

- *YIQ*

YIQ is nearly identical to the YUV colour space and composed of luminance Y and chrominance components I and Q. These components are represented as the second pair located on the axes of the diagram (Fig. 2). As such, I and Q denote various coordinate systems on the self-same plane. Thus, these components can be represented in the RGB values, where I is matched to range B, and Q is

matched to range G. This colour space can also be converted to the RGB format through linear transformations and is represented by the following expressions [72, 75], which are used to transform RGB to the YIQ model:

$$Y = 0.299R + 0.587G + 0.114B,$$

$$I = 0.595716R - 0.274453G - 0.321263B,$$

$$Q = 0.211456R - 0.522591G + 0.311134B.$$

- *HSV, HSI and HSL*

‘Perceptual colour spaces are popular samples in skin detection’. In these colour spaces, I, V and L represent luminance components, and H and S denote chrominance components. Three colour spaces separate the components’ saturation S, hue H and luminance components I, V and L. The colour spaces are distortions of the RGB colour cube and can be mapped from the RGB space by nonlinear transformation. Moreover, these colour spaces allow users to identify the boundary of the skin colour class intuitively in terms of hue and saturation, and this capability of colour spaces is considered to be the most important feature of these samples. ‘As I, V, or L provides the information of brightness, these components are usually dropped to mitigate the illumination dependency of skin colour’ [75–77].

- *IHLS*

IHLS is ‘also known as improved hue, luminance and saturation (IHLS) colour space’. It is enhanced with respect to identical colour spaces, such as HLS, HSI and HSV, through normalisation to remove the luminance component. Therefore, this method overcomes all difficulties limited by colour components, thereby providing a good distribution of the features of the space [78, 79].

- *CIEXYZ*

CIEXYZ is ‘one of the perceptual uniformity systems that show that a small perturbation to a component value is approximately equally perceptible across the range of a value’. The Commission International de l’Eclairage (CIE) colour system has been dependent on the CIE primaries since 1931. The CIEXYZ colour space forms a cone-shaped space with Y as the luminance component and X and Z as chrominance components. The luminance component of each colour space is dropped to form a 2D colour. Therefore, the values of each component of the colour spaces are adapted to the range of 0–255 and quantised in 256 levels [72, 80].

- *CIELAB*

CIELAB is ‘a reasonably perceptually uniform colour space proposed by the CIE’. This colour space has two

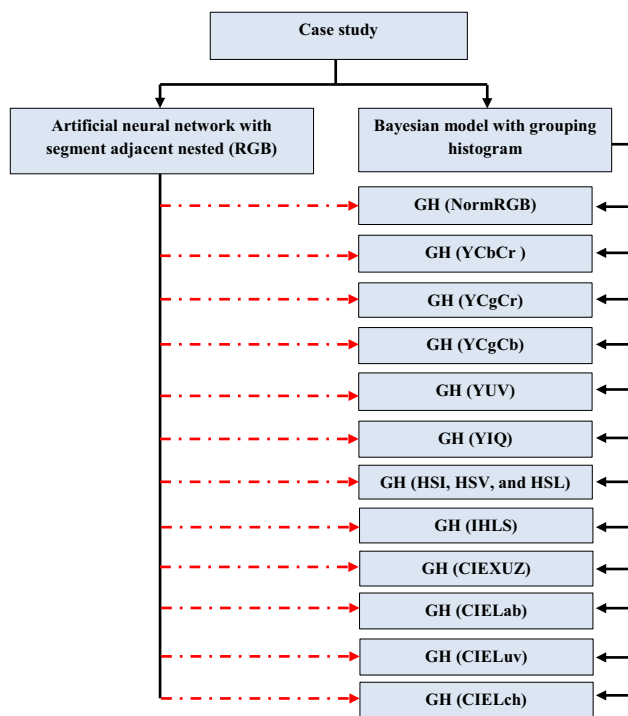


Fig. 2 Proposed skin detectors based on ANN methods involving SAN and a Bayesian model by using GH with different 12 colour spaces classes

**Table 1** List of parameters that were used for the three layers of the ANN-based skin detector

Name of parameter	Input layer	Hidden layer	Output layer
Number of nodes	9	4	1
Training function	trainlm (Levenberg–Marquardt algorithm)		
Number of epochs	1200		
Performance function	MSE (mean square error)		
Transfer function	Zero	Tansig (hyperbolic tangent sigmoid)	Purelin (pure linear)
Activation function	Deactivated	Nonlinear	Linear
Training goal	0.0001		

components, which are presented as a and b values according to the 1976 CIELAB colour space. Thus, a and b are the chrominance components in the colour space, and L is the luminance component. In general, CIELAB does not use the luminance component of the colour because it varies frequently across the human skin. Typically, chrominance is used to separate the skin from surrounding nonskin regions [81].

- *CIELUV*

CIELUV is ‘another colour space derived from a perceptually uniform colour space proposed by the CIE’. V and U are the chrominance components, and L is the luminance component. In general, the nonlinear transformation of CIELUV and CIELAB is applied to correct the problem that the RGB colour space is not perceptually uniform. Thus, the CIELUV colour space is a suitable sample when the luminance component is disregarded [69, 82].

- *CIELCH*

CIELCH is ‘a colour space that is also derived from perceptual uniformity systems created by CIE in which L represents luminance, and c and h denote the chrominance components’. Conducting Utans for the first time usually causes the illumination components to drop, resulting in multiple errors in the light cluster. Thus, lighting pixels depend on illumination components, and dark clusters (absence of light) perform well even when illumination components are omitted. The needed minimal components cannot be achieved by utilising Utans, possibly mitigating the features, training and testing on the network [83]. Thereafter, different colour spaces developed using AI models based on the literature are adopted in the present work. The total number of colour spaces is 14, which is divided into 12 classes, namely normalised RGB, YCbCr, YCgCr, YCgCb, YUV, YIQ, (HSI, HSV, HSL), IHLS, CIEXYZ, CIELAB, CIELUV and CIELCH. The reason is that the colour space class of HSI, HSL and HSV, where luminance element is deleted from each colour space (I, L and V) and the chroma element (HS), is retained, which is common between colour space groups. Thus, the

processing is applied only once. This process aims to acquire different alternatives (Fig. 2).

### 2.2.1.1.3 Training operation of a neural network model

In a previous study [54], the case study applies a neural network model in the proposed multi-agent learning of the skin detector that comprises three main layers, namely input, hidden and output layers. The architecture of this model comprises nine neurons in the input layer, four neurons in the hidden layer and a single neuron in the output layer. The model aims to conduct segmentation for different samples and is implemented using the RGB colour space, which is considered an essential vector for an image sample. The dataset is distributed into three parts, namely 1200 samples for training, 300 samples for validation and 300 samples for testing. A validation process is implemented in the set-up of the training process, which is conducted to calculate the mean square error (MSE) and represent the performance function of the ANN model.

The training operation for ANN adopts the back-propagation pattern, which is a suitable way to train the feed-forward neural network model. The training operation is achieved by 331,282,971 pixels based on 1200 images from the dataset of the skin and nonskin pixels. The default training function adopted a Levenberg–Marquardt approach in the back-propagation function (trainlm) [84]. The feed-forward network training aims to create a network object. The feed-forward operation requires five steps to generate a network object. In the first step, an array for input vectors is created to implement a segment adjacent-nested (SAN) technique. In the second step, the probability elements are identified as skin and nonskin indexes by creating an array that includes an output sample considered to be a target vector. In the third step, the 114 creations of an array identify the hidden layer size in the network. In the fourth step, the cell array, including the names of transfer functions used in two layers, is determined.

The default transfer function is used for hidden and output layers to achieve three layers of the network only. The hyperbolic tangent sigmoid (tansig) is utilised for the hidden layer, whilst the linear transfer function (purelin) is employed for the output layer. In the fifth step, the name of

the training function to be used is included [84]. By contrast, the activation function is used for three layers. The input layer is deactivated, and the hidden and output layers are nonlinear and linear, respectively. In the current study, the back-propagation feed-forward neural network is utilised on the basis of the reasonable results obtained by this method from previous studies [26, 85–87]. The RGB colour space is used only during the training of the ANN model. Table 1 describes the flow and characteristics of the neural part of the skin detector.

The training stage often takes a long time because of the very large amount of calculation that is required on the data. The maximum training time is 8 h, whilst 01:13:58 h is the minimum, based on trial and error.

In this process, the weight matrices between the input and the hidden and output layers are initialized with random values. After repeatedly presenting features of the input samples and desired targets, it compare the output with the desired outcome, followed by error measurement and weight adjustment until the correct output for every input is attained. Furthermore, the hidden layer neurons are estimated using an activation function that features the hyperbolic tangent sigmoid transfer function, whereas the output layer neuron is estimated using the activation function that features the linear transfer function. The training algorithm used is Levenberg–Marquardt back-propagation. To train the neural network for skin detection, a segment is extracted and entered as training input data into the ANN. The quality of the training sets that enters into the network determines how well the detector performs. The training phase of the skin detector is illustrated in Fig. 3.

Experimentally, the ANNs with a single hidden layer and with the parameters given in Table 1 have yielded high-accuracy results. This conclusion about the design was driven by the experiment results shown in Table 2, which are graphically depicted in Fig. 4. This figure shows the identification of the aligned trained data on the target, where the imaginary diagonal represents the target of the ideal data regression. The trained data regression is 0.96574, whilst the ideal state is 1. In other words, the results show regression analysis for a reducible ANN between two targets, namely 1 for skin and  $-1$  for non-skin. The results show that the relationship between the network outputs ( $Y$ ) and the targets ( $T$ ) is close and almost perfect; the correlation coefficient  $R$  is equal to 0.96574, which is almost an ideal fit. Moreover, from training these pixels, it can be concluded that the training pixels for both the skin and the nonskin using the segment adjacent-nested technique applied to the back-propagation neural network method do not have any overlap.

Figure 5 shows the error characteristic that progressively decreases until it reaches a stable stage. The error

rate is calculated using the MSE function, and the error decreases to  $10^{-4}$ , as shown in Table 1, over 1200 iterations

The final decision was to use the ANN as part of the proposed multi-agent learning of the skin detection, as mentioned in Table 1. Several separate training experiments and the results of their empirical tuning are given in Table 2. In the other words, Table 2 lists all of the trial experiments and the equivalent design for each experiment. These experiments were performed with different parameters, after which the best results were taken. This table shows the performance error minimisation, the linear regression of the targets relative to the outputs and other arguments that highlight the most efficient combination.

Table 2 concludes the following:

- Increasing the number of iterations is not crucial. This statement implies that sometimes we can increase the number of iterations but fail to notice a distinct decrease in the performance (error).
- It can be observed that there is significant improvement in the three-layer architecture compared with the equivalent two- and four-layer architectures, for the same parameters. In other words, the results show that the three-layer architecture is better than the other layered architectures for the same criteria.
- The training function ‘trainlm’ gives the best results compared to the ‘trainrp’ and ‘learnqdm’ with the same parameter selection.

**2.2.1.1.4 Training operation of the Bayesian model** The Bayesian model is considered important in machine learning algorithms and derived from the Bayesian rules and normalisation of the lookup table function (LUT). The procedure of the Bayesian model is based on clustering histograms by using 1200 images to calculate the LUT to identify skin and nonskin pixels. Histogram computation is typically implemented in post-training followed by the normalisation of LUT, thereby providing the distribution of separate probabilities. The training process of the Bayesian model is illustrated in Fig. 6.

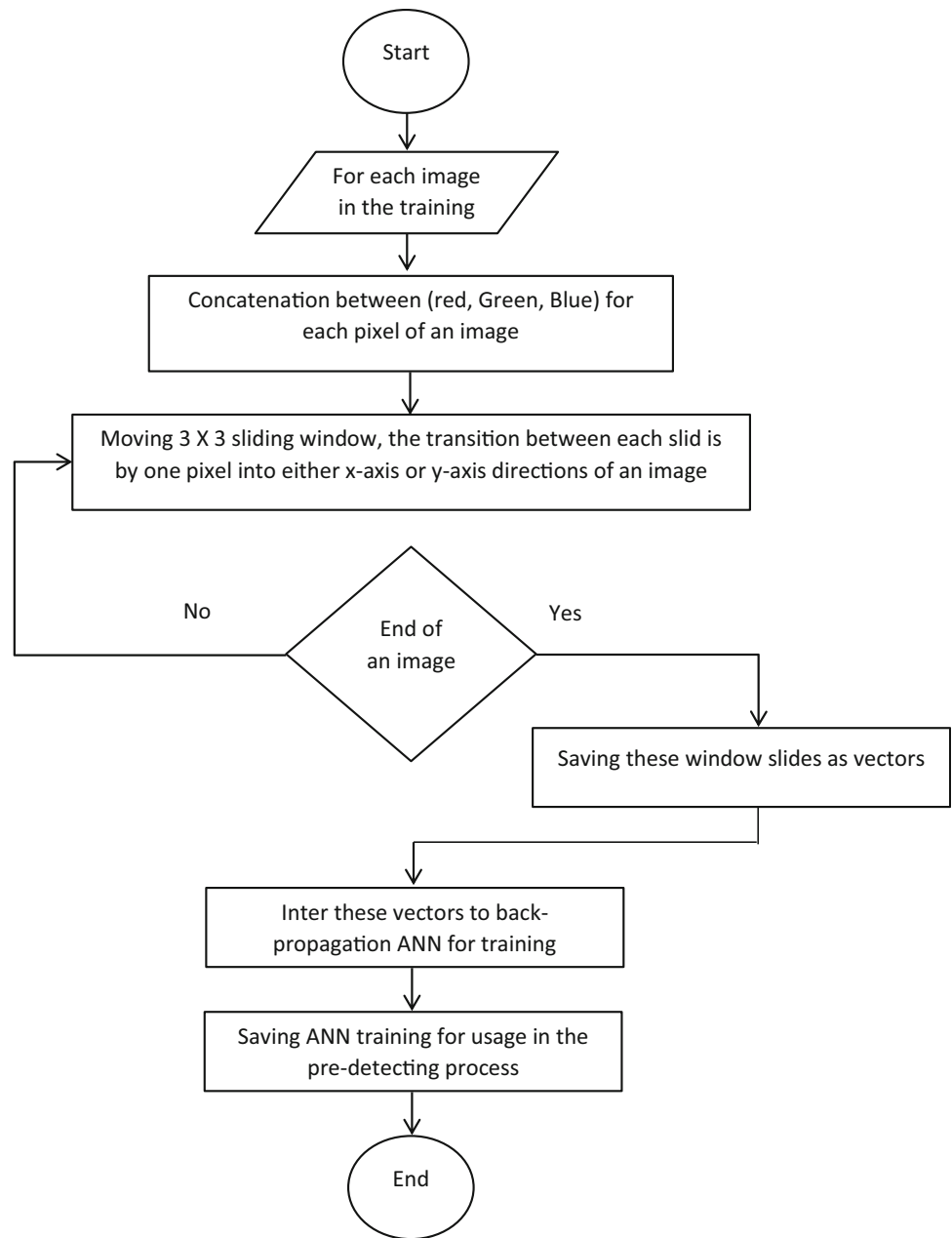
The probability of different colour space pixels calculated as a pixel of colour space (AdaptColorSpace) for identifying the skin pixel is observed. This probability is denoted by  $P_{\text{skin}}(\text{AdaptColorSpace})$  and represented by Eq. 1.

$$P_{\text{skin}}(\text{AdaptColorSpace}) = \frac{\text{Skin}[\text{AdaptColorSpace}]}{\text{Norm}}, \quad (1)$$

where  $\text{skin}[\text{AdaptColorSpace}]$  refers to the histogram value that matches the colour vector (AdaptColorSpace). The calculation of grouping histogram values by using the normalisation coefficient is represented in the parameter



**Fig. 3** Training phase of the proposed skin detector in ANN



Norm, which is a summation of all grouping histogram values. The normalisation value of LUT refers to the colour-matching probabilities of the skin. Thus, skin detection can be calculated as  $P(\text{skin}|\text{AdaptColorSpace})$  by following the Bayesian rule, which is given in Eq. 2 [70, 88]:

$$P(\text{Skin} \setminus \text{AdaptColorSpace}) = \frac{P(\text{Skin} \setminus \text{AdaptColorSpace}) \cdot P(\text{Skin})}{P(\text{Skin} \setminus \text{AdaptColorSpace}) \cdot P(\text{Skin}) + P(\sim \text{Skin} \setminus \text{AdaptColorSpace}) \cdot P(\sim \text{Skin})}, \quad (2)$$

The probability values of  $P(\text{Skin} \setminus \text{AdaptColorSpace})$  and  $P(\sim \text{Skin} \setminus \text{AdaptColorSpace})$  are directly calculated for skin and nonskin pixels by using grouping histogram

processes. Conversely, the previous probabilities, such as  $P(\text{skin})$  and  $P(\sim \text{skin})$ , can be easily computed by calculating the total number of pixels of the skin and nonskin at the training step (Eqs. 3, 4, respectively) [70]:

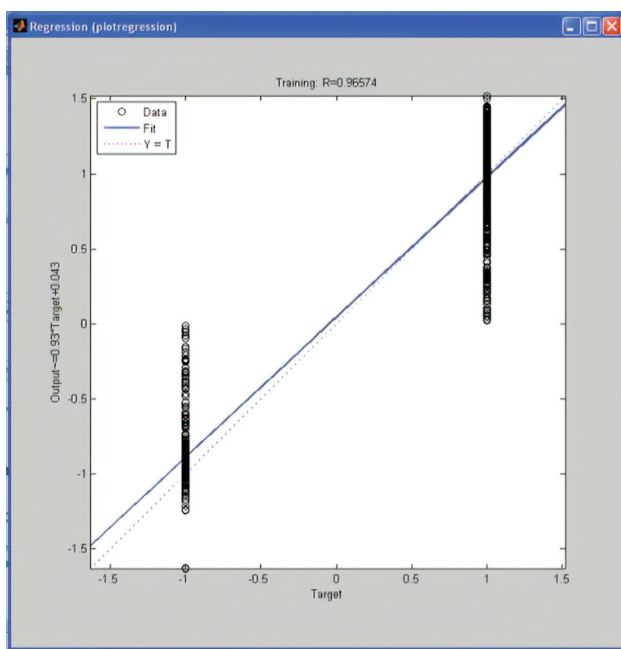
$$P(\text{Skin}) = \frac{T_s}{T_s + T_n}, \quad (3)$$

$$P(\sim \text{Skin}) = \frac{T_n}{T_n + T_s}, \quad (4)$$

where  $T_s$  and  $T_n$  are the values of skin and nonskin pixels, respectively.

**Table 2** List of all of the training experiments used in the ANN-based skin detector

Training experiments	Number of epochs	Performance	Regression	Gradient	Number of layers	Number of nodes	Training
Run1	1500	0.002	0.95	0.001	3	[9,2,1]	trainlm
Run2	1300	0	0.962	0	3	[9,4,1]	trainlm
Run3	1200	0	0.965	0	3	[9,4,1]	trainlm
Run4	1100	0	0.964	0	3	[9,4,1]	trainlm
Run5	1000	0.004	0.943	0.002	3	[9,6,1]	trainlm
Run6	800	0.006	0.941	0.005	3	[9,8,1]	trainlm
Run7	1200	0.356	0.936	0.192	3	[9,4,1]	trainrp
Run8	1200	0.738	0.926	0.203	3	[9,4,1]	learnngdm
Run9	1200	1.912	0.911	0.385	4	[9,8,6,1]	trainlm
Run10	1200	3.21	0.905	0.507	2	[9,1]	trainlm

**Fig. 4** Regression analysis for reducible ANN-based skin detector between two targets, i.e.  $-1$  for nonskin pixels and  $1$  for skin pixels

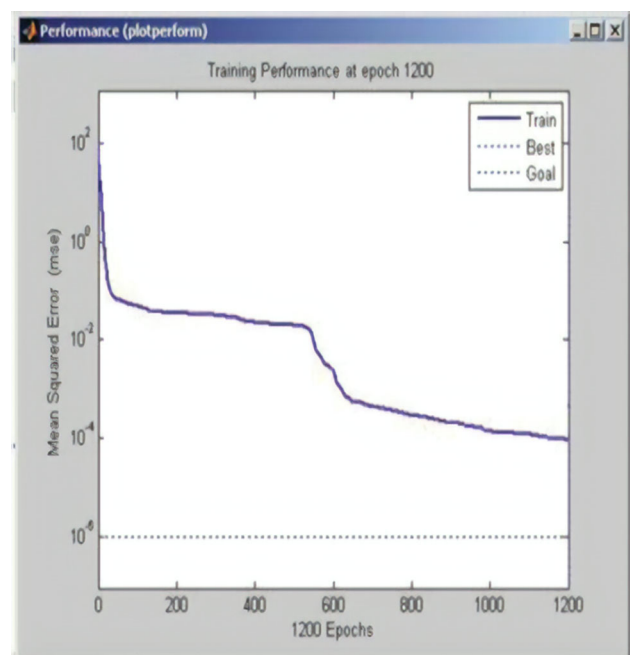
The results of the probability values should be saved within two files after the training process of the dataset is finished. These values can be represented as

$$Ps = \text{FUN}(\text{GH}(X)) \text{ and}$$

$$Pns = \text{FUN}(\text{GH}(Y)).$$

Thus, a multi-agent learning technique has been implemented on the basis of the desired goal based on specific functions to achieve the best results.

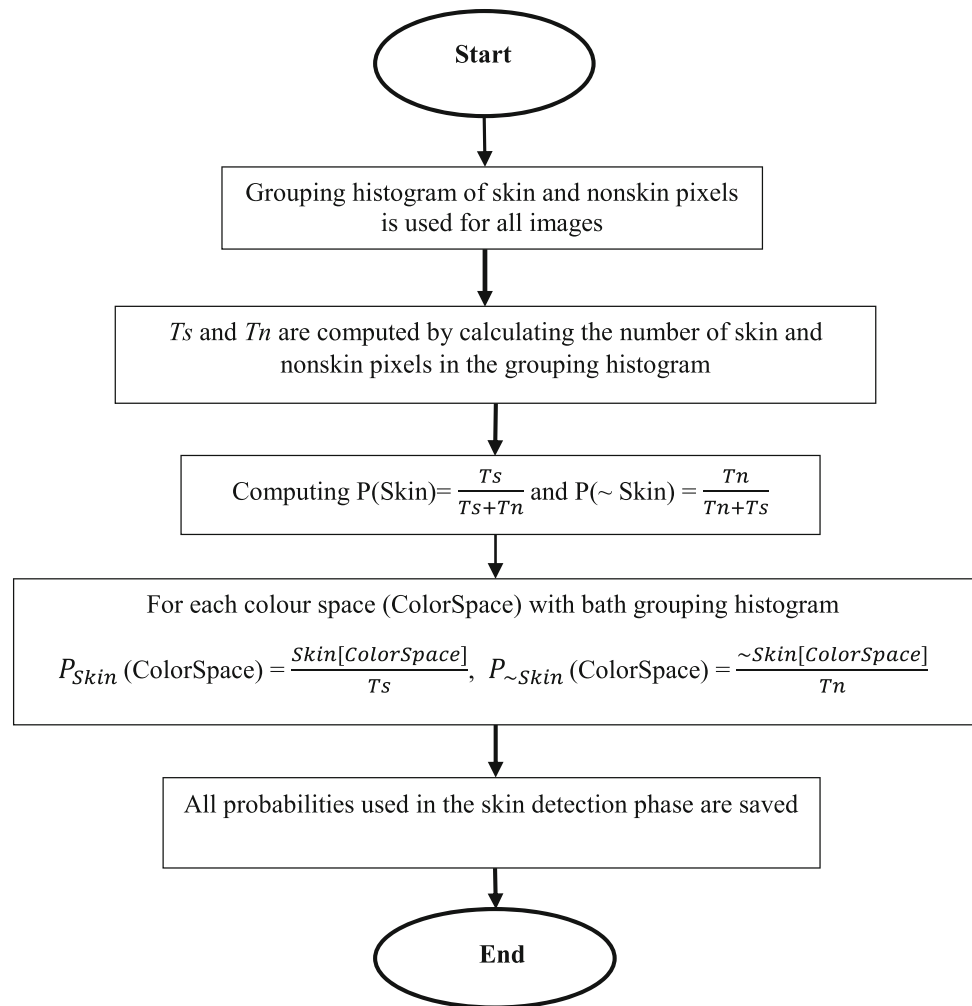
**2.2.1.1.5 Detection of the skin detector** The detection phase begins after the training phase is completed to gather the required data from various image samples by using the proposed technique [54].

**Fig. 5** Training error rates using the ANN-based skin detector

This phase evaluates the performance of the multi-agent learning technique adapted using different colour spaces. Figure 7 illustrates the process of segmentation and skin detection.

The detection process depends on the parameters obtained from the training processes of ANN and Bayesian. Therefore, these parameters are determined on the basis of the final stage of the proposed system based on the output of the original image. The system generates a new image that represents the skin pixels only, whereas a white background corresponds to the nonskin pixels. The original image pixels correspond to the inputs or probabilities represented in the two files ( $Ps$  and  $Pns$ ), which represent a pre-detection phase in the Bayesian model based on the

**Fig. 6** Training process of the Bayesian model [57]



aforementioned system. Thus, these pixels are saved in two variables:

$nn = P(\text{AdaptColorSpace} \setminus \text{skin}) \rightarrow$  skin pixels

$ww = P(\text{AdaptColorSpace} \setminus \sim \text{skin}) \rightarrow$  nonskin pixels

If  $P(ww)$  and  $P(\text{ColorSpace} \setminus ww)$  are represented as ( $ww = \{\text{skin or nonskin}\}$ ), then  $P(ww \setminus \text{ColorSpace})$ , which is already an accepted result that allows the application of the rule of the Bayesian model, is determined:

$$\text{IF } \frac{P(\text{AdaptColorSpace} \setminus \text{skin})}{P(\text{AdaptColorSpace} \setminus \sim \text{skin})} > \frac{P(\sim \text{skin})}{P(\text{skin})}, \quad (5)$$

where  $(\text{AdaptColorSpace})$  is classified as a skin pixel. Otherwise, it is a nonskin pixel.

Thus, the Bayesian rules can be computed at the minimum cost given [70]:

$$\frac{P(\text{AdaptColorSpace} \setminus \text{skin})}{P(\text{AdaptColorSpace} \setminus \sim \text{skin})} > \theta \rightarrow \text{AdaptColorSpace} \in \text{Skin}, \quad (6)$$

$$\frac{P(\text{AdaptColorSpace} \setminus \text{skin})}{P(\text{AdaptColorSpace} \setminus \sim \text{skin})} < \theta \rightarrow \text{AdaptColorSpace} \in \sim \text{Skin}. \quad (7)$$

According to Eq. 7, the calculation is such that Bayesian  $(pix_{x,y}) = ww/nn$ ; therefore, if  $nn$  is equal to 0, then  $nn$  automatically resets to  $nn = 0.00000000001$ . If  $ww = 0$ , then  $ww$  automatically resets to  $ww = 0.00000000001$ .

The threshold variable ( $\theta$ ) is defined in Eq. 8 as follows:

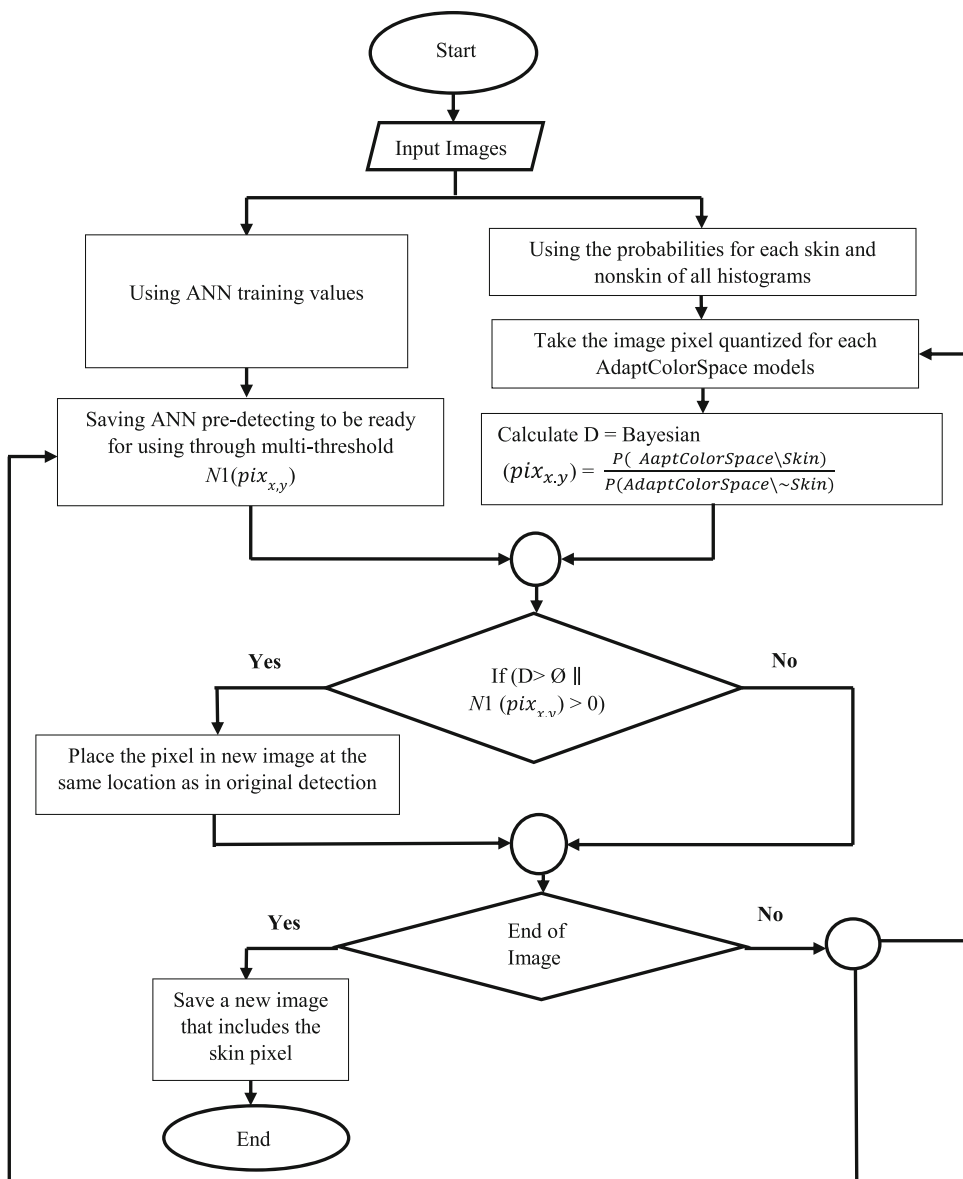
$$\theta = \frac{P(\sim \text{skin})}{P(\text{skin})} \quad (8)$$

The outcome of the multi-agent technique is distributed on the image pixels of  $I$  with the threshold value based on Eq. 9, which states that procedures are applied under the same conditions for both models.

For the neural network:

$$N1(pix_{x,y}) = \begin{cases} \sim \text{skin} & \text{if } N1(pix_{x,y}) < 0 \\ \text{skin} & \text{elsewhere} \end{cases}$$

**Fig. 7** Segmentation and detection of skin pixels



For Bayesian:

$$\text{Bayesian}(pix_{x,y}) = \begin{cases} \sim skin & \text{if } \text{Bayesian}(pix_{x,y}) < \theta \\ skin & \text{elsewhere} \end{cases}$$

$$\forall Pix_{x,y}, N1(Pix_{x,y}) < 0 \text{ OR } \text{Bayesian}(Pix_{x,y}) < \theta \rightarrow Pix_{x,y} \in \sim skin$$

(9)

where  $N1(Pix_{x,y})$  is the second parameter for function FUN ( $I, NI, Ps, Pns$ ). Bayesian ( $Pix_{x,y}$ ) is collected from the previous steps. The thresholds are considered on the basis of two aspects. Firstly, ( $N1(Pix_{x,y}) < 0$ ) represents the neural part, where 0 is selected based on Fig. 4. It was shown that the range of training for nonskin pixels was lower than zero, and the range of training for skin pixels

was greater than zero. From these points, the boundary between the skin and nonskin pixels is considered 0. Secondly, the procedure of the Bayesian model is represented in Bayesian ( $pix_{x,y}) < \theta$ . Nine threshold values ( $\theta$ ), namely 0.5, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9 and 0.95, are selected based on the study by Zaidan et al. [54].

Accordingly, the nine experiments for each colour space individually (as mentioned in Sect. 2.2.1.1.2) will be implemented in order to present 108 algorithms. In summary, the conditions are applied to the pixels; if the conditions are satisfied, then the pixels refer to the nonskin and [255 255 255] is assigned to refer to the white pixel. Otherwise, [R (y, x), G (y, x), B (y, x)] is returned from the original image I. The collected pixels from the set

**Table 3** Establishment of the decision matrix

Algorithm	Criteria			
	Reliability	Time complexity	Error rate for (training and validation)	
Algorithm 1	RV (A1/TS)	TcV (A1/TS)	ERT (A1/TS <sub>t</sub> )	ERV (A1/VS)
Algorithm 2	RV (A2/TS)	TcV (A2/TS)	ERT (A2/TS <sub>t</sub> )	ERV (A2/VS)
Algorithm 3	RV (A3/TS)	TcV (A3/TS)	ERT (A3/TS <sub>t</sub> )	ERV (A3/VS)
Algorithm 4	RV (A4/TS)	TcV (A4/TS)	ERT (A4/TS <sub>t</sub> )	ERV (A4/VS)
Algorithm 5	RV (A5/TS)	TcV (A5/TS)	ERT (A5/TS <sub>t</sub> )	ERV (A5/VS)
...	...	...	...	...
Algorithm 108	RV (A <sub>n</sub> /TS)	TcV (A <sub>n</sub> /TS)	ERT (A <sub>n</sub> /TS <sub>t</sub> )	ERV (A <sub>n</sub> /VS)

*RV* reliability values, *TcV* time complexity values, *ERT* error rate for training, *ERV* error rate for validation, *A* algorithm, *TS* test samples, *n* number of algorithms, *TS<sub>t</sub>* training samples, *VS* Validation samples

condition are skin pixels. Thus, the algorithms collect skin pixels in accordance with the conditions.

**2.2.1.2 Crossing between the developed skin detector and three main groups of criteria** Table 3 shows the decision matrix that established based on the crossover between 108 algorithms and three main groups of criteria are collected from the literature. The procedure highlights the effect of the different criteria on various colour spaces. Therefore, the obtained multiple criteria include the basic elements in the established decision matrix.

In summary, the crossover between multiple criteria and adopted colour spaces creates the decision matrix, which is used to generate the final results in the next phase.

Table 3 shows the three main groups of criteria with 108 algorithms used as alternatives in the proposed decision matrix. The procedures for each criterion are discussed in detail as follows.

**2.2.1.2.1 Computation of reliability group elements** The reliability group involves three basic sections, namely parameter matrix, relationship and behaviour. The relationship emphasises the importance of evaluating the skin detectors in our study through these sections [32, 56].

The procedure for each sub-criterion within a reliability group is explained in detail.

The parameter matrix comprises four key parameters, namely TP, TN, FN and FP, which are the backbone for computing the remaining criteria in the reliability group [89–91].

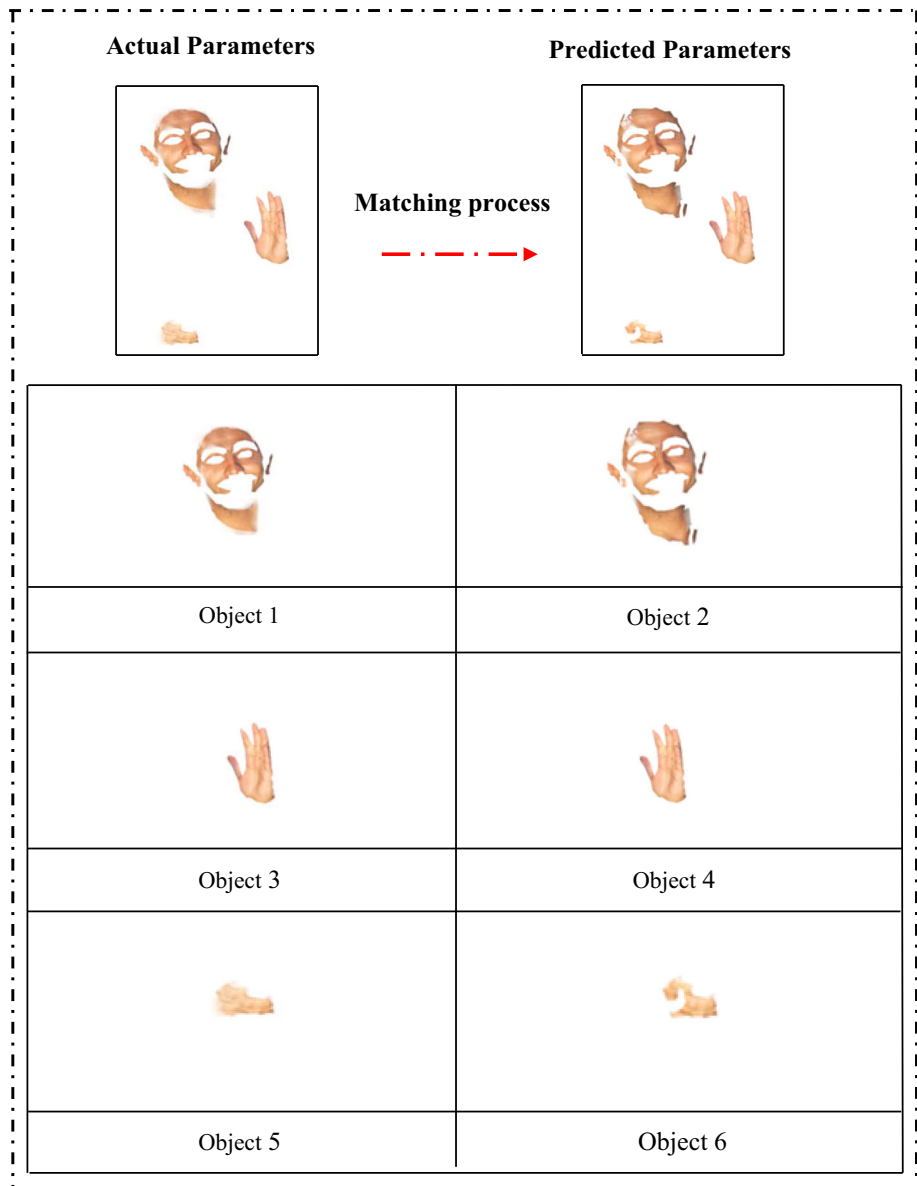
These parameters introduce the results obtained by a previously implemented multi-agent learning technique. A certain procedure is performed to calculate the basic parameters and their integral values based on the matching process. Figure 8 presents additional details about the matching process for a sample of the predicted parameters and the actual parameters used to create the confusion matrix.

The object locations of each image are matched to calculate the image's skin pixels and to compute the locations of pairs of images according to their object location. For example, the objects from 1 to 6 in Fig. 5 represent the base of both images. Therefore, the number of pixels in the first object is computed with the second object by a pointer, which matches each pixel of the actual parameters with every pixel of the predicted parameters. This pointer computes the skin pixels of the actual parameters by only using the predicted parameters, which present a standard measure for the number of pixels of the predicted parameters (computed as TP). The differences between the standard and calculated pixels are presented as FN. The remaining pixels are computed for the other objects. For example, the average value of the skin pixel is computed to yield the final TP for the entire image and attain the average FN as the definitive FN. Furthermore, the TN that represents the background of the image can be computed as nonskin pixels, whereas FP is considered a complement of TN. Thus, the parameter values are computed to be TN based on the values of the nonskin pixel objects from the predicted parameters, whereas FP is computed on the basis of the differences between the standard and calculated values.

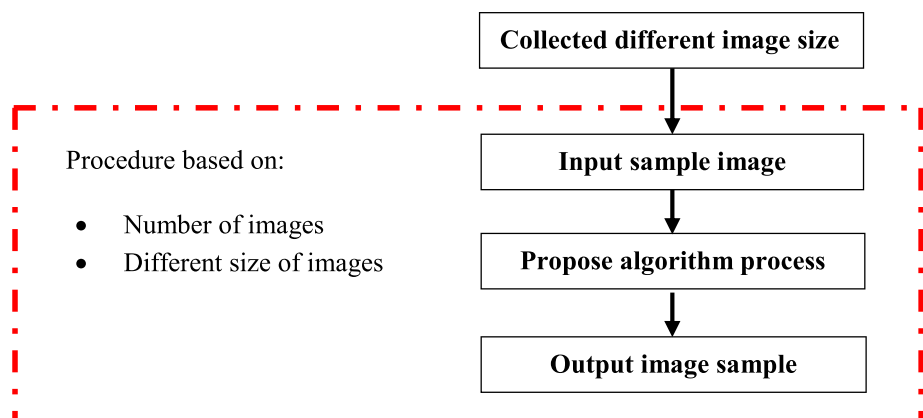
By contrast, these parameters are calculated on the basis of different threshold values for each colour space to attain the final result of the DM. As such, the values of 108 algorithms are separately calculated by conducting individual experiments to generate the final parameter values for the decision matrix.

**2.2.1.2.2 Computation of time complexity criterion** The time complexity criterion is important in this study. The procedure and methodology for calculating time depend on the time consumption of the output and input sample images. The flowchart (Fig. 9) shows the process of computing time complexity [20, 92].

**Fig. 8** Matching process for different objects



**Fig. 9** Procedure of time complexity



The image calculation processes rely on the number and size of the image samples.

$$T_{\text{process}} = T_o - T_i, \tag{10}$$

where  $T_o$  is the output time image process and  $T_i$  is the input time image process.

$$T_{\text{total}} = \frac{T_{\text{process}}}{T_{\text{average}}}, \tag{11}$$

where  $T_{\text{process}}$  represents the difference between the output and input image samples and  $T_{\text{average}}$  denotes the average processing time for all samples. These equations cover the computation time of different image sizes and various skin objects in the self-same image size.

**2.2.1.2.3 Error rate computation within dataset elements** According to the evaluation of several studies, the error rate within the dataset is the major property, which relies on machine learning algorithms. Special mechanisms in most AI models are used in these algorithms to produce results through the training and testing stages. Training is a key step in these algorithms. Therefore, the dataset is trained several times to derive a minimum error rate by selecting a particular dataset, including training and validation. However, reliability values in the testing data are addressed to gain the final outcomes. The training process is individually performed for 108 algorithms [90, 92].

**2.2.1.3 Evaluation and testing of developed skin detector according to three criteria groups** The evaluated and tested decision matrix depends on the calculation the criteria procedure for 108 algorithms. This calculation is performed from nine experiments according to the threshold values for data collection, thereby providing the final results of the decision matrix.

**2.2.2 Performance of decision matrix**

This step is considered as the second part of the identification and performance phase. This part is conducted in two trends: Firstly, the relationship between the criteria is investigated to calculate the correlation between the criteria in order to make sure, either needs to use all of the criteria in decision matrix. Secondly, performance analysis is conducted to evaluate and compare the criteria and identify the factors that affect their behaviour in order to made sure that use all the criteria in evaluation as multi-dimensional measurements or not.

**2.2.2.1 Correlation between criteria** This step mainly aims to investigate the relationship between the criteria and determine their degree of correlation. The case study includes three main groups of criteria that have

interconnected physical characteristics. Therefore, the relationship between these criteria must be proven. Software and techniques based on mathematical and statistical methods exist to prove the relationships between variables. A Pearson method is adopted to find a correlation between the various criteria used in our study [59, 60, 65].

Thus, the method that represented Pearson’s  $r$  is shown as follows:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X}) * (Y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 * (\sum_{i=1}^n (Y_i - \bar{y})^2)}, \tag{12}$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \tag{13}$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \tag{14}$$

where  $n$  is the number of input values,  $X$  and  $Y$  are two criteria,  $x$  and  $y$  are their average values and  $r$  is the correlation coefficient.

$r$  ranges from  $-1$  to  $1$ .  $r$  near  $0.00$  indicates uncorrelated criteria, whilst  $r$  near or equal to  $1$  indicates a high correlation level [66, 67].

Therefore, Pearson’s formula is applied to compute the correlation coefficient between different criteria. Multiple criteria are evaluated and tested on the basis of their features. In this study, various criteria that influence one another are independently collected. Therefore, determining the relationship between criteria and verifying the degree of correlation between them are imperative. Figure 10 shows the taxonomy of the criterion distribution in the three main layers.

Figure 10 illustrates the taxonomy constructed from the three layers, which comprises three major sets of criteria in our study. The first layer includes reliability criterion ( $R$ ), time complexity criterion ( $T_c$ ) and error rate within the dataset criterion (ER) groups. The second layer consists of three key sections, such as parameter matrix, relationship and behaviour, which are derived from the reliability criterion. The validation and training criteria are derived from the error ratio criterion. The third layer comprises 10 criteria. Amongst the 10 criteria, four are called confusion matrices, namely true positive (TP), true negative (TN), false positive (FP) and false negative (FN), which are derived from the matrix of parameters. The four other criteria are accuracy (ACC), precision (PR), recall (RE) and specificity (SP), which are derived from the relationship of parameters. The final two criteria are  $F$ -measure ( $F$ ) and  $G$ -measure ( $G$ ), which are derived from the behaviour of the parameters.

**2.2.2.2 Performance analysis of criteria** In order to make sure that utilise all the criteria in the evaluation process as

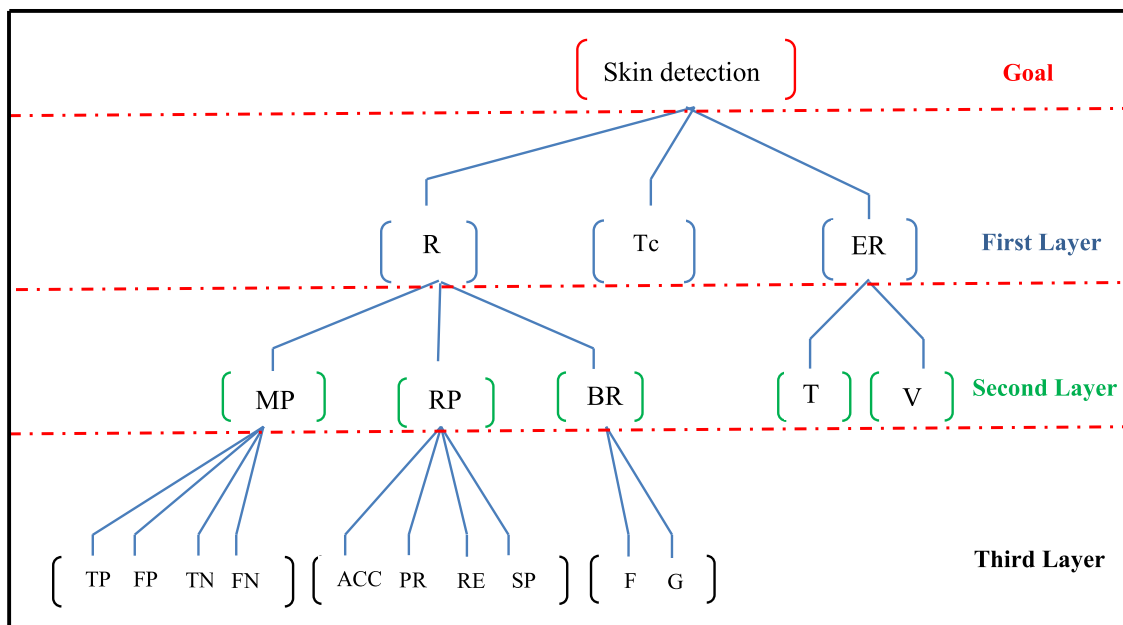


Fig. 10 Taxonomy of criteria distribution in the three layers

multidimensional measurements or not, performance analysis is conducted to compare the criteria and identify the factors that affect their behaviour. Practically, a total of criteria that resulted from Sect. 2.2.2.1 should be evaluated to determine their behaviour on the basis of 108 algorithms. To show the differences between the criteria behaviour in different scenarios, the threshold values for each criterion should be distributed to the different 12 colour spaces used. Thus, the performance analysis of the criteria implemented for the three main groups of criteria used in this study is reliability group, time complexity group and error rate within the dataset group.

### 2.3 Development phase

The development phase involves the designing of a new methodology based on MCDM techniques. Many MCDM techniques or methods, which use different concepts, such as simple additive weighting (SAW), weighted sum model (WSM), weighted product method (WPM), multiplicative exponential weighting (MEW), multi-objective optimisation methods, hierarchical adaptive weighting (HAW), analytic hierarchy process (ANP), AHP and TOPSIS [93–102], have been investigated. None of the aforementioned methods have been used to evaluate and benchmark real-time IoT skin detection approaches. Figure 11 explains the techniques, drawbacks and recommendations of popular MCDM techniques in accordance with previous studies [32, 37, 56, 88, 95, 103–120].

Ultimately, a new research direction in the application of MCDM techniques is created by combining two or more techniques to avoid the deficiencies from relying on a single technique [121–123]. Extensively, the integration of TOPSIS and AHP has become acceptable within MCDM techniques for the following reasons: calculation of the relative distance by using weights and objective data; presentation of results in full ranking; provision of appropriate random analysis and prevention of trade-off through the use of nonlinear relations, which enable conversion to a programmatic procedure [124–126]. Many studies have provided rankings for alternatives and arranged the problems [21, 125, 127–129]. According to the literature, two or more techniques are commonly integrated to address various problems in many fields except skin detection applications. Therefore, a methodological approach must be identified to resolve this gap.

In summary, integrating the AHP to identify each weight for all criteria based on experts' preferences with TOPSIS is recommended to present the total ranking of alternatives for real-time IoT skin detection approaches. Accordingly, the proposed methodology relies on the integration between TOPSIS and AHP in ranking and selecting the best alternatives. This method is based on the values of the decision matrix created in the previous phase [98]. The implementation of this phase is discussed in the next paragraphs. Figure 12 describes the new methodology for evaluating and benchmarking real-time IoT skin detection approaches.



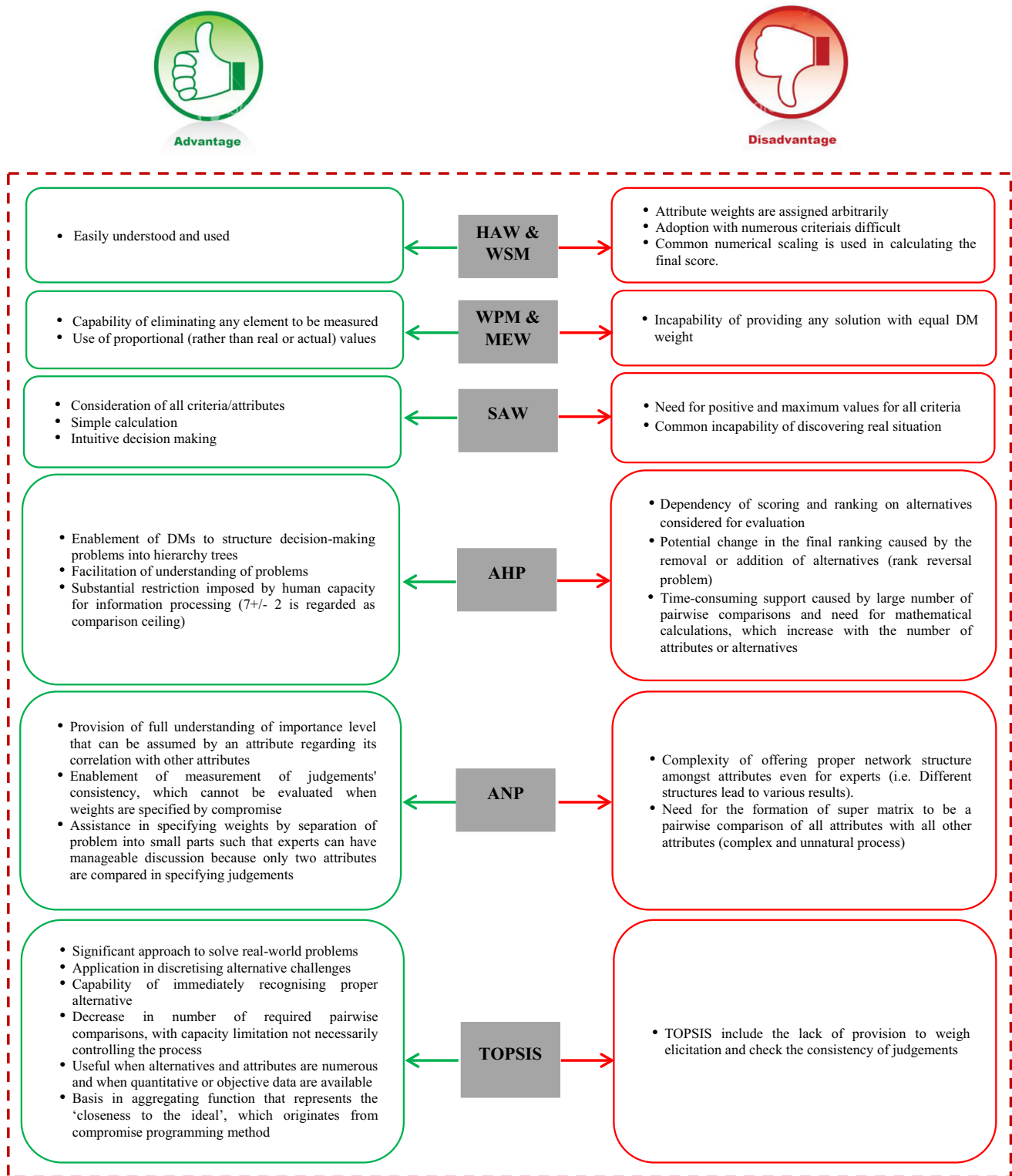
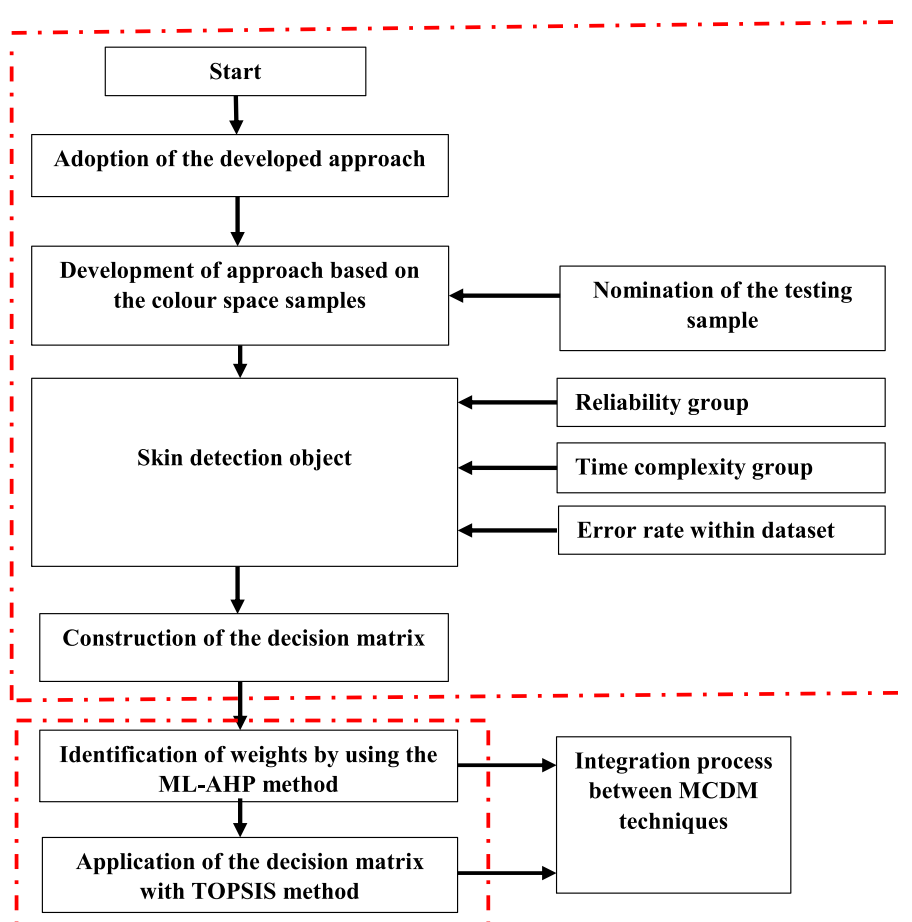


Fig. 11 Advantages and limitations of MCDM methods

**Fig. 12** New methodology for evaluating and benchmarking real-time IoT skin detector



### 2.3.1 Developing a decision-making solution for the skin detection approach based on integrated ML-AHP and TOPSIS

The integration of ML-AHP and TOPSIS is widely accepted by many researchers for the following reasons. Firstly, these methods can present the results of complete ranking and calculate the relative distance based on weights and objective data. Secondly, their results are satisfactory for random analyses and show a harmonious trade-off by using nonlinear relationships, thereby allowing easy conversion into a programmable format. Thus, in our study, ML-AHP and TOPSIS are integrated to rank numerous colour space algorithms for real-time IoT skin detection approaches.

A total of 13 weight settings that represent the three key groups of criteria under different circumstances are used in the first part. In this step, the weights are assigned according to external evaluator preferences. Thus, the weights can be measured through AHP in a pairwise form, and the outcome of this technique is subsequently used in the TOPSIS method. Different colour spaces are

developed as alternatives in the decision matrix. These alternatives must be ranked to configure the selection of the best alternative in the second part. The TOPSIS method adopts the decision matrix to provide the final results.

### 2.3.2 Adoption of ML-AHP to investigate the weights of different evaluators

Skin detection approaches are being developed to achieve certain objectives (e.g. skin and nonskin detection in images). For each evaluation criterion, developers can assign the weight according to these objectives. The definition of weights depends on the priority preference in MCDM with interval numerals [119, 123, 130]. On the issue of skin detection evaluation and benchmarking, assigning weights is difficult. By contrast and in normal life problems, experts can help assign weights. Differences in expert opinions on the importance of skin detection criteria can increase these difficulties, thereby creating a conflict with the designer's objective. Weights can be assigned in multiple manners and achieved by the ML-AHP algorithm to conduct

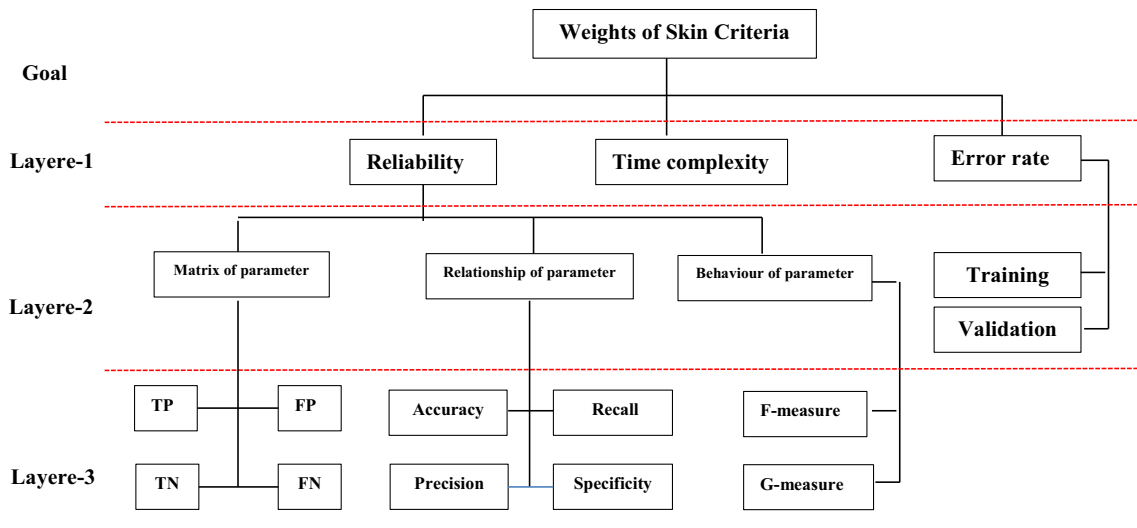


Fig. 13 Weights of criteria in a multi-layer structure

pairwise comparisons amongst the criteria and adopt weights to settle these problems [131]. Moreover, the solutions of the aforementioned problems depend on experts, thereby creating a conflict between the preference of experts and that of designers. A total of 13 different weight settings are selected and assigned in the decision-making process to solve this problem. Thus, our research uses a pairwise technique (i.e. ML-AHP method). Figure 13 shows the AHP method based on multiple layers.

**2.3.2.1 Pairwise comparisons for each criterion** AHP has been proposed as a technique to obtain ratio scales from paired comparisons and is now a widely known MCDM method [117, 132].

ML-AHP allows a few inconsistencies in judgement because of human imprecision. The ratio scales are derived from principal eigenvectors, and the consistency index is derived from the principal eigenvalue. The following equation represents the required number of pairwise comparisons:

$$n * (n - 1) / 2, \tag{15}$$

where  $n$  represents the number of attributes utilised during evaluation. In comparing a set of criteria,  $n$  in pairs is based on the amount of relative weights. Criteria and weights can be represented as  $(C_1 \dots C_n)$  and  $(w_1 \dots w_n)$ , respectively.

Table 4 Sample pairwise comparison matrix

Criteria	A	B	C
A	1	A/B	A/C
B	B/A	1	B/C
C	C/A	C/B	1

This comparison can be represented in the matrix as follows:

$$C = \begin{matrix} & \begin{matrix} C_1 & C_2 & \dots & C_n \end{matrix} \\ \begin{matrix} C_1 \\ \vdots \\ C_2 \\ \vdots \\ C_n \end{matrix} & \begin{matrix} w_1/w_1 & w_1/w_2 & \dots & w_1/w_n \\ w_2/w_1 & w_2/w_2 & \dots & w_2/w_n \\ \vdots & \vdots & \ddots & \vdots \\ w_n/w_1 & w_n/w_2 & \dots & w_n/w_n \end{matrix} \end{matrix}$$

The matrix uses a pairwise ratio whose rows provide the ratio of the weights for each element with respect to all other ratios. This method focuses on extracting the weights of various activities based on importance. Typically, importance is a judgement based on different criteria. Occasionally, these criteria correspond to objectives selected by activities for investigation [132].

Table 4 presents a comparison matrix for priority rating, which comprises three pairwise elements in the decision matrix.

In this study, six experts who are from various universities in Malaysia and have sufficient background in image processing involving AI methods are invited to participate as evaluators. They are asked to compare different criteria based on the priorities identified in accordance with their perspectives. Their responses are gathered, adopted and attached with the degree of consistency according to the rules of hierarchy theory to calculate the weights. Figure 14 shows the formula of the questionnaire presented to the evaluators.

Reference [133] proposed a new scale to calculate the degree of importance amongst different criteria. The scale uses the difference between successive scale values to

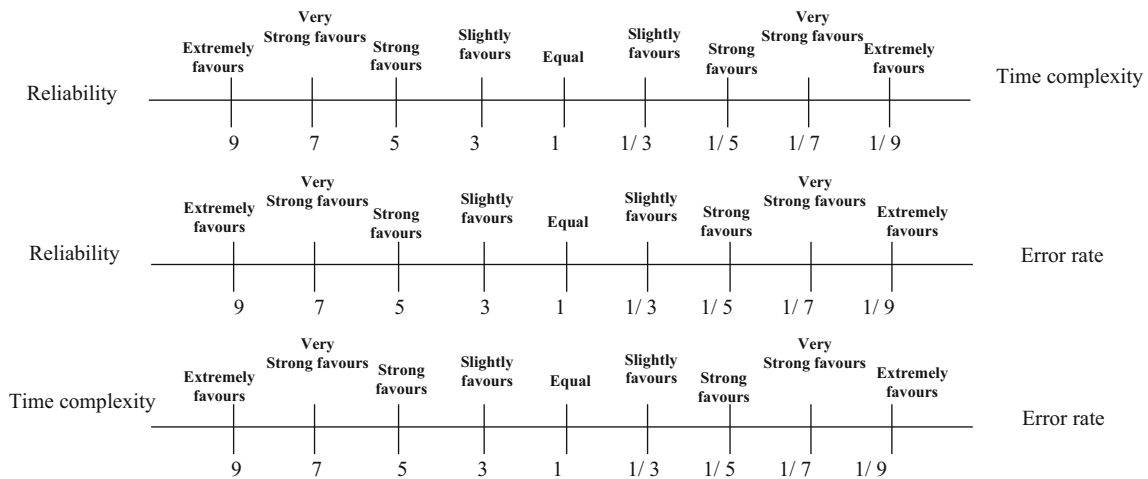


Fig. 14 Pairwise answer from evaluators

Table 5 Intensity scale of criteria

Degree of importance	Description
1	Equal importance
3	Weak importance of one over another
5	Essential or strong importance
7	Demonstrated importance
9	Absolute importance
2, 4, 6, 8	Intermediate values between the two adjacent judgements

compare the criteria within the scale values that range from 1 to 9. Table 5 shows an initial step towards the construction of the intensity scale of importance for activities [43].

**2.3.2.2 Design of ML-AHP measurement structure** In this step, our work includes multiple layers to distribute the criteria. The ML-AHP measurement matrix is implemented to attain the weights based on the preferences of the evaluator. ML-AHP measurements use mathematical calculations that are based on pairwise comparisons to convert the experts’ judgements into weights for each criterion. A CR must also be calculated for judgements that represent the internal consistency values entered. Thereafter, the answers of the evaluators with pairwise comparisons are collected, and the reciprocal matrix is created. The reciprocal matrix provides the sub-criteria values for each level’s main criteria and identifies the importance of each feature compared with its parent. Thus, the obtained features of the main criteria represent the importance of each feature in relation to the goal. Figure 15 shows the weights computed through ML-AHP measurement based on different evaluators.

**2.3.2.3 Calculating the weights of criteria and checking the consistency value** Various responses gathered from

different evaluators must be converted into numerical values in the decision matrix. For these values, the decision matrix implements procedures, such as normalisation and aggregation. In the next step, the weights of the criteria are determined and ranked. By contrast, the ML-AHP measurement considers an important vector to conduct a consistency test, which is normally required after the criteria weights are completely calculated. Inconsistency is often observed in the answers obtained by the ML-AHP questionnaire from individual evaluators, thereby affecting the overall consistency of the test. Hence, the CR must be tested before all responses are collected from the evaluators [134].

Finally, CR can be measured to determine the consistency of the pairwise comparison. This procedure is called a consistency index. A CR larger than 0.10 indicates inconsistency in the pairwise comparison, whereas a CR equal to or less than 0.10 indicates a reasonable comparison [98, 135].

The following formula calculates the CR:

$$CR = CI/RI \tag{16}$$

CI represents the consistency index, which is derived from the following:

$$CI = (\max - n)/(n - 1) \tag{17}$$

RI can be obtained from Table 6 as follows.

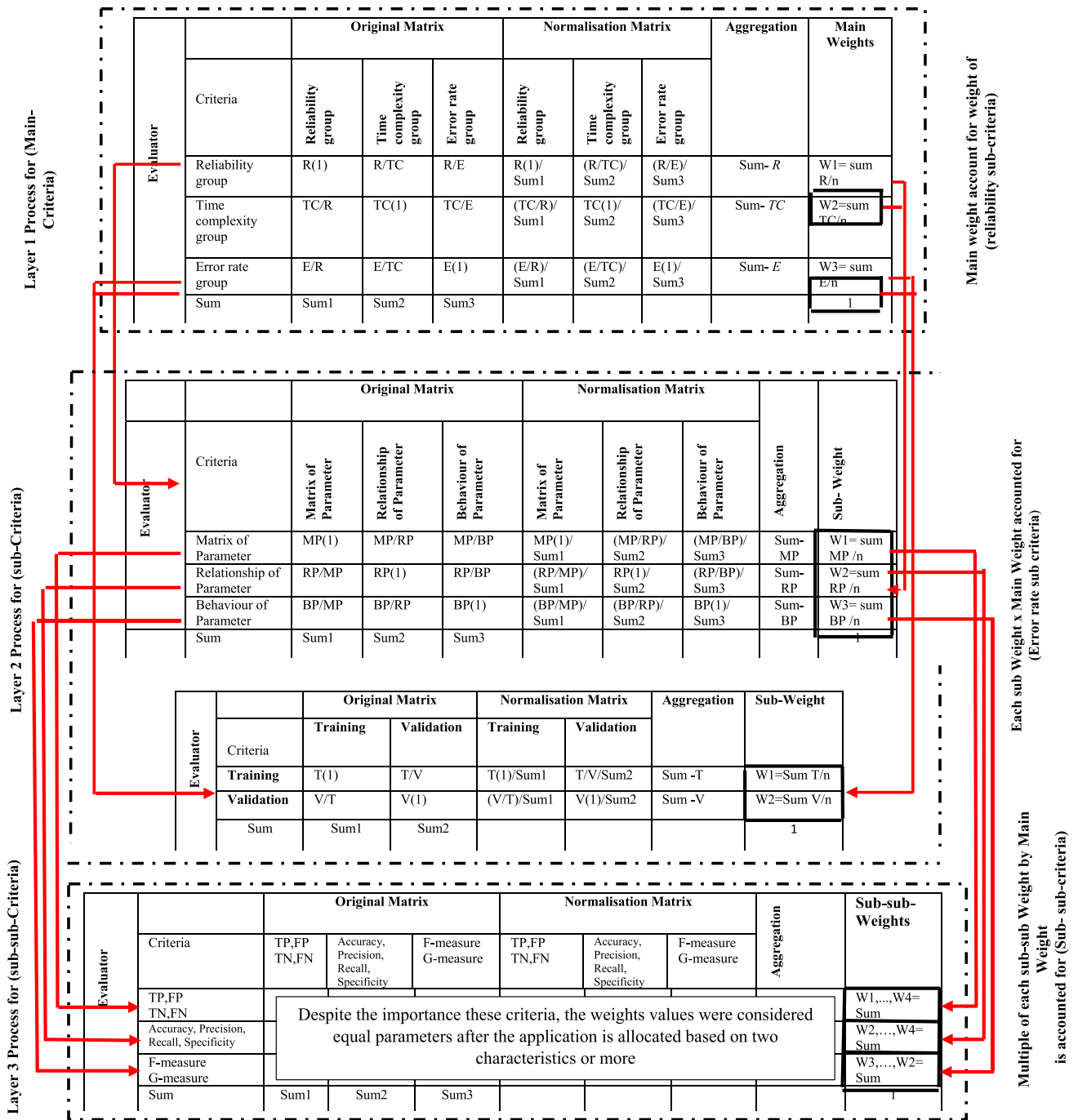


Fig. 15 Steps of ML-AHP used to account the matrix multiple layers

Table 6 Random index (RI) [117]

N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
RI	0.00	0.00	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49	1.51	1.54	1.56	1.58	1.59

### 2.3.3 Utilisation of TOPSIS for evaluation and benchmarking of real-time IoT skin detection approaches

TOPSIS, which is favoured amongst MCDM techniques, is utilised in this procedure, which involves several steps [136, 137]. TOPSIS is applied to all alternatives based on the geometric distance from the positive and negative ideal solutions [138]. Thus, the most suitable alternative is ‘the alternative with the longest geometric distance to the negative ideal solution and the shortest geometric distance to the positive ideal solution’ [97, 139].

According to [126], TOPSIS consists of six steps: (1) constructing the normalised decision matrix, (2) constructing the weighted normalised decision matrix, (3) identifying the ideal and negative ideal solutions, (4) separating measurement calculations based on the Euclidean distance, (5) identifying closeness to the ideal solution calculation and (6) ranking the alternatives based on the closeness to the ideal solution. Each step is explained in detail as follows.

#### Step 1

In this step, the different criterion dimensions are transformed into a nondimensional criterion, and the comparison of each attribute is permitted.  $(x_{ij})_{m \times n}$  matrix is then normalised from  $(x_{ij})_{m \times n}$  to  $R = (r_{ij})_{m \times n}$  by the normalisation method.

$$r_{ij} = x_{ij} / \sqrt{\sum_{i=1}^m x_{ij}^2} \tag{18}$$

This step results in a new matrix  $R$  as follows:

$$R = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ r_{m1} & r_{m2} & \dots & r_{mn} \end{bmatrix} \tag{19}$$

#### Step 2

$w = w_1, w_2, w_3, \dots, w_j, \dots, w_n$  represents the group of weights from the decision-maker, which provides the normalised decision matrix. The new matrix is calculated by multiplying every column from the normalised matrix ( $R$ ) with related weight  $w_j$ . Notably, the summation of the weights should be equal to 1, which is expressed as follows:

$$\sum_{j=1}^m w_j = 1. \tag{20}$$

This step results in a new matrix  $V$ , which is expressed as follows:

$$V = \begin{bmatrix} v_{11} & v_{12} & \dots & v_{1n} \\ v_{21} & v_{22} & \dots & v_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ v_{m1} & v_{m2} & \dots & v_{mn} \\ w_1 r_{11} & w_2 r_{12} & \dots & w_n r_{1n} \\ w_1 r_{21} & w_2 r_{22} & \dots & w_n r_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ w_1 r_{m1} & w_2 r_{m2} & \dots & w_n r_{mn} \end{bmatrix} \tag{21}$$

#### Step 3

The definition of the two artificial alternatives, namely  $A^*$  and  $A^-$ , is explained in detail.  $A^*$  represents the ‘ideal alternative’, and  $A^-$  represents the ‘negative ideal alternative’.

$$A^* = \left\{ \left( \left( \max_i v_{ij} | j \in J \right), \left( \min_i v_{ij} | j \in J^- \right) | i = 1, 2, \dots, m \right) \right\} \tag{22}$$

$$= \{v_1^*, v_2^*, \dots, v_j^*, \dots, v_n^*\} \tag{23}$$

$$A^- = \left\{ \left( \left( \min_i v_{ij} | j \in J \right), \left( \max_i v_{ij} | j \in J^- \right) | i = 1, 2, \dots, m \right) \right\} \tag{24}$$

$$= \{v_1^-, v_2^-, \dots, v_j^-, \dots, v_n^-\} \tag{25}$$

where  $J$  the benefit attribute and subset of  $\{i = 1, 2, \dots, m\}$ , whereas  $J^-$  is the set of cost attribute and complement set of  $J$  or  $(J^c)$ .

#### Step 4

In this step, the separation of measurements is calculated by computing the distance between every alternative within  $V$  and the ideal vector  $A^*$  on the basis of the Euclidean distance as follows:

$$S_{i^*} = \sqrt{\sum_{j=1}^n (v_{ij} - v_j^*)^2}, \quad i = (1, 2, \dots, m). \tag{26}$$

The separation for every alternative within  $V$  from  $A^-$ , which is a negative ideal, is computed as follows:

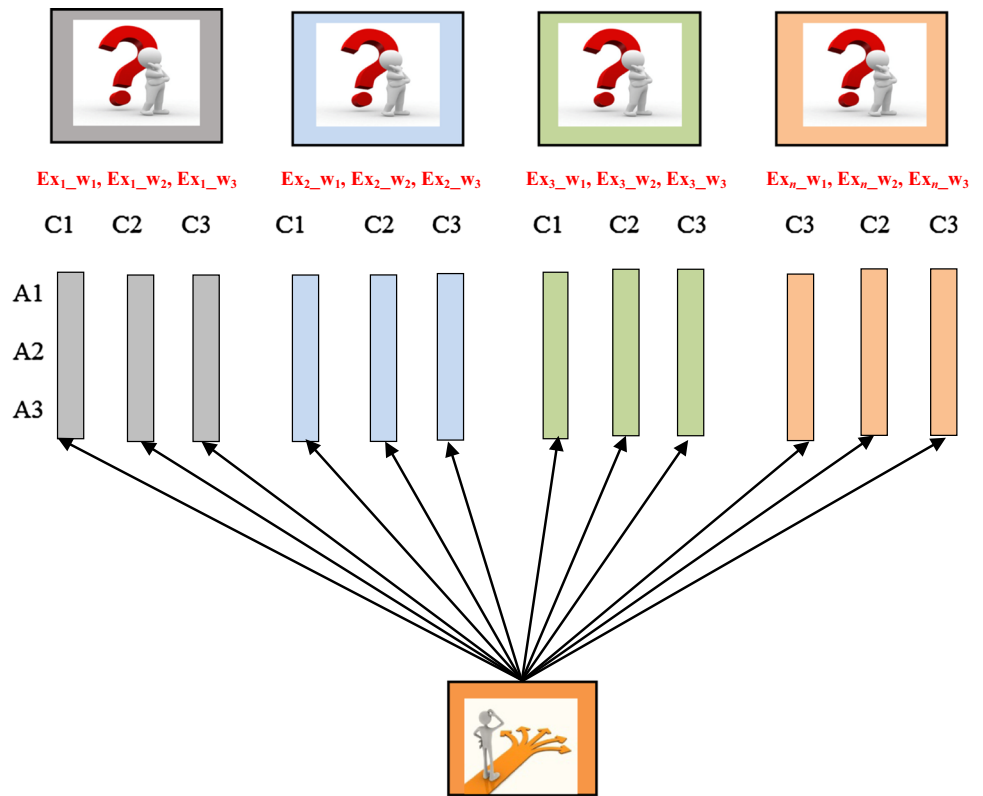
$$S_{i^-} = \sqrt{\sum_{j=1}^n (v_{ij} - v_j^-)^2}, \quad i = (1, 2, \dots, m). \tag{27}$$

This step counts the alternatives ( $S_{i^*}$  and  $S_{i^-}$ ) that represent the distance between every alternative and between the negative ideal and ideal alternatives.

#### Step 5

In this process, the closeness of  $A_i$  to the ideal solution  $A^*$  can be calculated as follows:

Fig. 16 GDM process



$$C_i^* = S_{i^-} / (S_{i^-} + S_{i^*}), \quad 0 \leq C_i^* \leq 1, i = (1, 2, \dots, m). \tag{28}$$

$C_i^* = 1$  if and only if  $(A_i = A^*)$ . Similarly,  $C_i^* = 0$  if and only if  $(A_i = A^-)$ .

**Step 6**

Finally, a group of alternative  $A_i$  can be prioritised on the basis of  $C_i^*$ , where the lowest value means the worst performance and high values refer to a remarkable performance.

**2.3.3.1 Decision-making context** Couple fundamental decision-making contexts are emphasised on the basis of the individual and group decision-makers (GDMs). This situation is encountered by individuals when they collectively select amongst given alternatives. Afterwards, the decision can no longer be attributed to any individual member of the group because all the individuals within such social group processes as social influence contribute to the outcome. GDM techniques systematically gather and integrate the awareness for the judgement of experts from various areas [112, 140–142]. Thus, the GDMs of experts or evaluators subjectively provide their judgement and weights for the criteria. Figure 16 shows the group decision-making process where  $Ex_1, Ex_2, \dots, Ex_n$  are the GDMs of experts, which can provide their judgement and weights for the criteria.

For example, with the problems of real-time IoT skin detection evaluation and benchmarking considered in the context of group decision-making, reliability (C1),

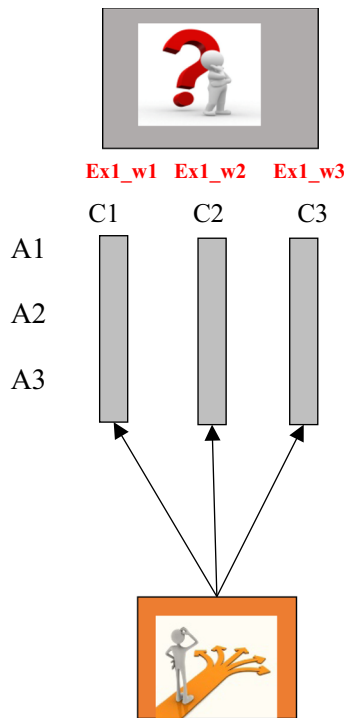
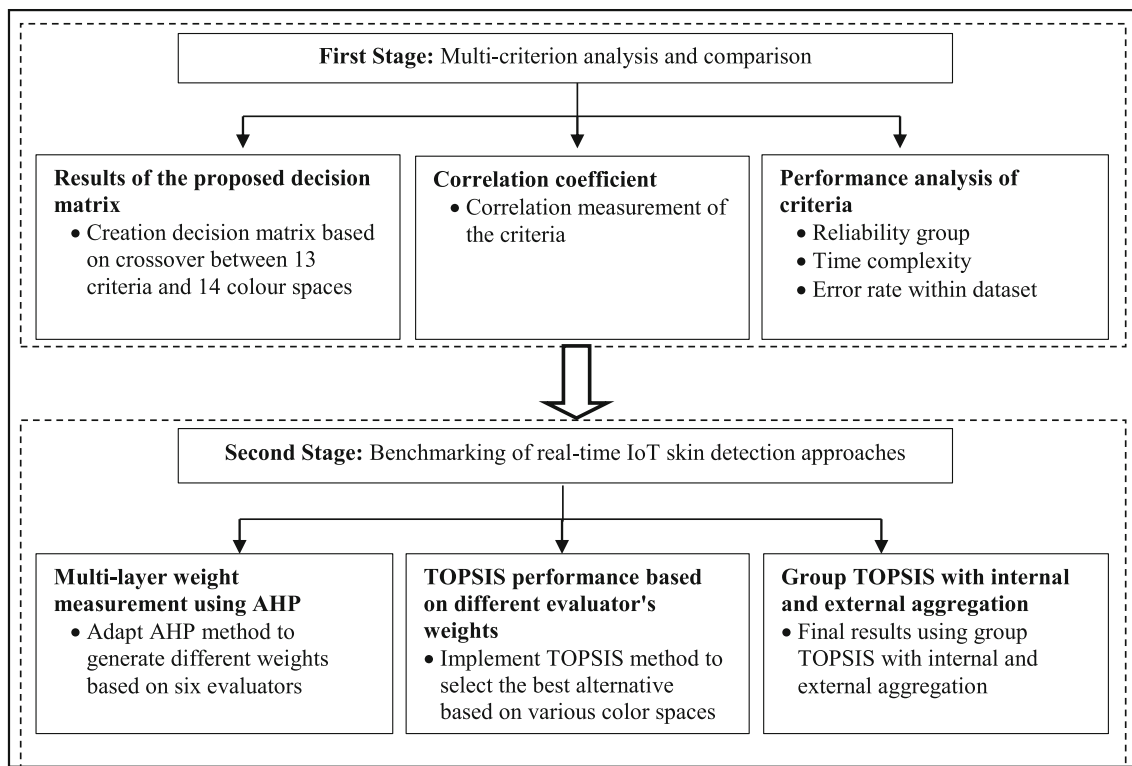


Fig. 17 Individual decision-maker process



**Fig. 18** Structure of the result

time complexity (C2) and error rate within the dataset (C3) are subjectively measured by the evaluators (Fig. 17).

Two different settings in the original context of decision-making, which fundamentally depend on the problem itself, are considered. Normal decision-making is a first setting where an individual provides subjective judgements and weights for all criteria. Another setting is group decision-making, where decision-makers provide their subjective judgements and their own weights for every criterion as a group. In this study, the problems of evaluation and benchmarking in real-time IoT skin detection approaches for all the data are objectively obtained (numerical values). Thus, the context of the decision-making process must be considered.

### 3 Result

This section presents the result based on two stages, namely analysis and comparison of multi-criteria and benchmarking of real-time IoT skin detection approaches (Fig. 18). In the first stage, the results of the proposed decision matrix, correlation coefficient and performance analysis of criteria are presented. In the second stage, the results of multi-layer weight measurement, TOPSIS performance based on different evaluator's weights and group

TOPSIS with internal and external aggregations are presented.

#### 3.1 Multi-criteria analysis and comparison

In this section, the results of proposed decision matrix, correlation and performance analysis between criteria have been presented in the following subsections.

##### 3.1.1 Results of the proposed decision matrix

In this section, the results obtained based on the development of a case study were adapted. A total of 14 colour spaces were developed on the basis of multi-agent technique using two artificial intelligence models. The outcome of the development process generated parameters that considered fundamental values to calculate the values of the reliability group. In addition, the other basic criteria values were calculated according to their methodology. Thus, the multiple criteria values with skin detection engine values produced the decision matrix that represents the dataset in our study.

The decision matrix was constructed using the values of the 13 criteria and 14 colour spaces. The values of each colour space were calculated individually based on the nine threshold values. Thus, the criteria values were calculated according to the threshold values that generated nine



**Table 7** Implemented decision matrix

Colour Spaces	TN	TP	FP	FN	Accuracy	Recall	Precision	Specificity	F-measure	G-measure	$T_c$ (s)	ERV	ERT
Normalised RGB	79.5	80.06	20.5	19.94	79.9003	0.8006	0.7961	0.7950	0.7984	0.7984	8.10	0.00003	0.00004
	81.32	80.22	18.68	19.78	81.7211	0.8022	0.8111	0.8132	0.8066	0.8066	8.14	0.00003	0.00004
	82.04	79.56	17.96	20.44	82.4378	0.7956	0.8158	0.8204	0.8056	0.8057	8.01	0.00003	0.00004
	83.77	80.91	16.23	19.09	84.1746	0.8091	0.8329	0.8377	0.8208	0.8209	8.08	0.00003	0.00004
	85.84	80.12	14.16	19.88	86.2406	0.8012	0.8498	0.8584	0.8248	0.8251	8.25	0.00003	0.00004
	90.45	81.3	9.55	18.7	90.8565	0.813	0.8949	0.9045	0.852	0.853	8.31	0.00003	0.00004
	91.84	79.99	8.16	20.01	92.24	0.7999	0.9074	0.9184	0.8503	0.852	8.32	0.00003	0.00004
	93.72	78.62	6.28	21.38	94.1131	0.7862	0.926	0.9372	0.8504	0.8533	8.41	0.00003	0.00004
YCbCr	92.91	79.14	7.09	20.86	93.3057	0.7914	0.9178	0.9291	0.8499	0.8522	8.35	0.00003	0.00004
	80.22	81.92	19.78	18.08	80.6296	0.8192	0.8055	0.8022	0.8123	0.8123	9.82	0.0001	0.00007
	85.27	84.37	14.73	15.63	85.6919	0.8437	0.8514	0.8527	0.8475	0.8475	9.84	0.0001	0.00007
	86.49	85.78	13.51	14.22	86.9189	0.8578	0.8639	0.8649	0.8609	0.8609	9.86	0.0001	0.00007
	88.95	86.61	11.05	13.39	89.3831	0.8661	0.8869	0.8895	0.8764	0.8764	9.91	0.0001	0.00007
	91.64	89.27	8.36	10.73	92.0864	0.8927	0.9144	0.9164	0.9034	0.9035	9.95	0.0001	0.00007
	91.9	90.85	8.1	9.15	92.3543	0.9085	0.9181	0.919	0.9133	0.9133	9.85	0.0001	0.00007
	96.55	94.53	3.45	5.47	97.0227	0.9453	0.9648	0.9655	0.9549	0.955	9.96	0.0001	0.00007
YCrCb	99.55	98.68	0.45	1.32	100.043	0.9868	0.9955	0.9955	0.9911	0.9911	9.9	0.0001	0.00007
	95.3	93.94	4.7	6.06	95.7697	0.9394	0.9524	0.953	0.9458	0.9459	9.94	0.0001	0.00007
	81.72	82.6	18.28	17.4	82.133	0.826	0.8188	0.8172	0.8224	0.8224	11.61	0.00015	0.0001
	84.64	85.12	15.36	14.88	85.0656	0.8512	0.8471	0.8464	0.8492	0.8492	11.63	0.00015	0.0001
	85.84	86.47	14.16	13.53	86.2724	0.8647	0.8593	0.8584	0.862	0.862	11.64	0.00015	0.0001
	86.62	88.84	13.38	11.16	87.0642	0.8884	0.8691	0.8662	0.8786	0.8787	11.62	0.00015	0.0001
	89.47	91.81	10.53	8.19	89.9291	0.9181	0.8971	0.8947	0.9075	0.9075	11.71	0.00015	0.0001
	92.56	90.81	7.44	9.19	93.0141	0.9081	0.9243	0.9256	0.9161	0.9162	11.73	0.00015	0.0001
YCbCr	93.59	97.38	6.41	2.62	94.0769	0.9738	0.9382	0.9359	0.9557	0.9559	11.66	0.00015	0.0001
	98.63	99.66	1.37	0.34	99.1283	0.9966	0.9864	0.9863	0.9915	0.9915	11.74	0.00015	0.0001
	93.49	96.59	6.51	3.41	93.973	0.9659	0.9369	0.9349	0.9512	0.9513	11.71	0.00015	0.0001
	82.59	81.87	17.41	18.13	82.9994	0.8187	0.8246	0.8259	0.8217	0.8217	11.94	0.00016	0.00015
	85.08	84.73	14.92	15.27	85.5037	0.8473	0.8503	0.8508	0.8488	0.8488	11.92	0.00016	0.00015
	86.41	85.89	13.59	14.11	86.8395	0.8589	0.8634	0.8641	0.8611	0.8611	11.91	0.00016	0.00015
	88.74	86.68	11.26	13.32	89.1734	0.8668	0.885	0.8874	0.8758	0.8759	11.9	0.00016	0.00015
	91.79	89.53	8.21	10.47	92.2377	0.8953	0.916	0.9179	0.9055	0.9056	11.94	0.00016	0.00015
YCbCr	90.84	92.64	9.16	7.36	91.3032	0.9264	0.91	0.9084	0.9181	0.9182	11.92	0.00016	0.00015
	97.25	93.65	2.75	6.35	97.7183	0.9365	0.9715	0.9725	0.9537	0.9538	11.92	0.00016	0.00015
	99.61	98.67	0.39	1.33	100.103	0.9867	0.9961	0.9961	0.9914	0.9914	11.94	0.00016	0.00015
	96.52	93.52	3.48	6.48	96.9876	0.9352	0.9641	0.9652	0.9494	0.9496	11.93	0.00016	0.00015
	82.84	81.46	17.16	18.54	83.2473	0.8146	0.826	0.8284	0.8203	0.8203	12.02	0.00018	0.00013
	85.72	84.14	14.28	15.86	86.1407	0.8414	0.8549	0.8572	0.8481	0.8481	12.19	0.00018	0.00013
	86.68	85.37	13.32	14.63	87.1069	0.8537	0.865	0.8668	0.8593	0.8593	12.09	0.00018	0.00013
	87.58	87.73	12.42	12.27	88.0187	0.8773	0.876	0.8758	0.8766	0.8766	12.23	0.00018	0.00013
YUV	89.54	90.61	10.46	9.39	89.9931	0.9061	0.8965	0.8954	0.9013	0.9013	12.25	0.00018	0.00013
	93.27	89.62	6.73	10.38	93.7181	0.8962	0.9302	0.9327	0.9129	0.913	12.35	0.00018	0.00013
	93.62	96.8	6.38	3.2	94.104	0.968	0.9382	0.9362	0.9528	0.953	12.08	0.00018	0.00013
	96.81	98.9	3.19	1.1	97.3045	0.989	0.9688	0.9681	0.9788	0.9788	12.31	0.00018	0.00013
	94.04	95.42	5.96	4.58	94.5171	0.9542	0.9412	0.9404	0.9477	0.9477	12.42	0.00018	0.00013
	82.51	80.72	17.49	19.28	82.9136	0.8072	0.8219	0.8251	0.8145	0.8145	12.81	0.00019	0.00015
	85.02	84.71	14.98	15.29	85.4436	0.8471	0.8497	0.8502	0.8484	0.8484	12.83	0.00019	0.00015
	84.27	85.38	15.73	14.62	84.6969	0.8538	0.8444	0.8427	0.8491	0.8491	12.85	0.00019	0.00015

Table 7 (continued)

Colour Spaces	TN	TP	FP	FN	Accuracy	Recall	Precision	Specificity	F-measure	G-measure	$T_c$ (s)	ERV	ERT
	86.31	86.71	13.69	13.29	86.7436	0.8671	0.8636	0.8631	0.8654	0.8654	12.86	0.00019	0.00015
	91.48	88.36	8.52	11.64	91.9218	0.8836	0.9121	0.9148	0.8976	0.8977	12.86	0.00019	0.00015
	88.71	92.01	11.29	7.99	89.1701	0.9201	0.8907	0.8871	0.9052	0.9053	12.88	0.00019	0.00015
	97.82	91.38	2.18	8.62	98.2769	0.9138	0.9767	0.9782	0.9442	0.9447	12.92	0.00019	0.00015
	99.93	95.85	0.07	4.15	100.409	0.9585	0.9993	0.9993	0.9785	0.9787	12.95	0.00019	0.00015
	96.37	92.26	3.63	7.74	96.8313	0.9226	0.9621	0.9637	0.942	0.9422	12.95	0.00019	0.00015
HSI, HSV, HSL	80.55	81.11	19.45	18.89	80.9556	0.8111	0.8066	0.8055	0.8088	0.8088	8.51	0.00005	0.00006
	82.37	81.27	17.63	18.73	82.7764	0.8127	0.8217	0.8237	0.8172	0.8172	8.6	0.00005	0.00006
	83.09	80.61	16.91	19.39	83.4931	0.8061	0.8266	0.8309	0.8162	0.8163	8.53	0.00005	0.00006
	84.82	81.96	15.18	18.04	85.2298	0.8196	0.8437	0.8482	0.8315	0.8316	8.73	0.00005	0.00006
	86.89	81.17	13.11	18.83	87.2959	0.8117	0.8609	0.8689	0.8356	0.836	8.7	0.00005	0.00006
	91.5	82.35	8.5	17.65	91.9118	0.8235	0.9064	0.915	0.863	0.864	8.77	0.00005	0.00006
	92.8	79.63	7.2	20.37	93.1982	0.7963	0.9171	0.928	0.8524	0.8546	8.52	0.00005	0.00006
	92.64	76.56	7.36	23.44	93.0228	0.7656	0.9123	0.9264	0.8325	0.8357	8.55	0.00005	0.00006
	93.95	75.37	6.05	24.63	94.3269	0.7537	0.9257	0.9395	0.8309	0.8353	8.72	0.00005	0.00006
IHLS	81.61	82.37	18.39	17.63	82.0219	0.8237	0.8175	0.8161	0.8206	0.8206	9.12	0.00011	0.00009
	83.75	83.43	16.25	16.57	84.1672	0.8343	0.837	0.8375	0.8356	0.8356	9.18	0.00011	0.00009
	84.92	84.45	15.08	15.55	85.3423	0.8445	0.8485	0.8492	0.8465	0.8465	9.12	0.00011	0.00009
	85.26	85.67	14.74	14.33	85.6884	0.8567	0.8532	0.8526	0.8549	0.8549	9.23	0.00011	0.00009
	87.47	85.28	12.53	14.72	87.8964	0.8528	0.8719	0.8747	0.8622	0.8623	9.31	0.00011	0.00009
	92.25	84.93	7.75	15.07	92.6747	0.8493	0.9164	0.9225	0.8816	0.8822	9.32	0.00011	0.00009
	93.44	87.73	6.56	12.27	93.8787	0.8773	0.9304	0.9344	0.9031	0.9035	9.4	0.00011	0.00009
	96.14	86.28	3.86	13.72	96.5714	0.8628	0.9572	0.9614	0.9075	0.9088	9.37	0.00011	0.00009
	94.35	87.34	5.65	12.66	94.7867	0.8734	0.9392	0.9435	0.9051	0.9057	9.19	0.00011	0.00009
CIEXYZ	82.54	81.89	17.46	18.11	82.9495	0.8189	0.8243	0.8254	0.8216	0.8216	10.33	0.00016	0.00013
	85.37	84.21	14.63	15.79	85.7911	0.8421	0.852	0.8537	0.847	0.847	10.38	0.00016	0.00013
	86.55	85.82	13.45	14.18	86.9791	0.8582	0.8645	0.8655	0.8613	0.8613	10.39	0.00016	0.00013
	88.97	86.58	11.03	13.42	89.4029	0.8658	0.887	0.8897	0.8763	0.8763	10.41	0.00016	0.00013
	91.73	89.22	8.27	10.78	92.1761	0.8922	0.9152	0.9173	0.9035	0.9036	10.43	0.00016	0.00013
	91.93	90.81	8.07	9.19	92.3841	0.9081	0.9184	0.9193	0.9132	0.9132	10.32	0.00016	0.00013
	96.63	94.32	3.37	5.68	97.1016	0.9432	0.9655	0.9663	0.9542	0.9543	10.37	0.00016	0.00013
	99.61	98.52	0.39	1.48	100.103	0.9852	0.9961	0.9961	0.9906	0.9906	10.44	0.00016	0.00013
	95.25	93.98	4.75	6.02	95.7199	0.9398	0.9519	0.9525	0.9458	0.9458	10.44	0.00016	0.00013
CIELAB	82.6	81.79	17.4	18.21	83.009	0.8179	0.8246	0.826	0.8212	0.8212	10.52	0.00017	0.00014
	85.25	84.41	14.75	15.59	85.6721	0.8441	0.8513	0.8525	0.8477	0.8477	10.55	0.00017	0.00014
	86.08	85.98	13.92	14.02	86.5099	0.8598	0.8607	0.8608	0.8602	0.8602	10.57	0.00017	0.00014
	88.91	86.37	11.09	13.63	89.3419	0.8637	0.8862	0.8891	0.8748	0.8749	10.54	0.00017	0.00014
	91.81	89.16	8.19	10.84	92.2558	0.8916	0.9159	0.9181	0.9036	0.9037	10.62	0.00017	0.00014
	90.9	92.49	9.1	7.51	91.3625	0.9249	0.9104	0.909	0.9176	0.9176	10.67	0.00017	0.00014
	97.43	93.36	2.57	6.64	97.8968	0.9336	0.9732	0.9743	0.953	0.9532	10.68	0.00017	0.00014
	99.57	98.57	0.43	1.43	100.063	0.9857	0.9957	0.9957	0.9907	0.9907	10.64	0.00017	0.00014
	96.38	93.63	3.62	6.37	96.8482	0.9363	0.9628	0.9638	0.9494	0.9494	10.66	0.00017	0.00014
CIELUV	81.66	82.64	18.34	17.36	82.0732	0.8264	0.8184	0.8166	0.8224	0.8224	10.14	0.00014	0.00011
	84.52	85.25	15.48	14.75	84.9463	0.8525	0.8463	0.8452	0.8494	0.8494	10.16	0.00014	0.00011
	85.99	86.56	14.01	13.44	86.4228	0.8656	0.8607	0.8599	0.8631	0.8631	10.15	0.00014	0.00011
	86.57	88.93	13.43	11.07	87.0147	0.8893	0.8688	0.8657	0.8789	0.879	10.13	0.00014	0.00011
	89.34	91.87	10.66	8.13	89.7994	0.9187	0.896	0.8934	0.9072	0.9073	10.25	0.00014	0.00011
	92.49	90.92	7.51	9.08	92.9446	0.9092	0.9237	0.9249	0.9164	0.9164	10.22	0.00014	0.00011

**Table 7** (continued)

Colour Spaces	TN	TP	FP	FN	Accuracy	Recall	Precision	Specificity	F-measure	G-measure	T <sub>c</sub> (s)	ERV	ERT
CIELCH	93.41	97.42	6.59	2.58	93.8971	0.9742	0.9366	0.9341	0.9551	0.9552	10.28	0.00014	0.00011
	98.59	99.71	1.41	0.29	99.0886	0.9971	0.9861	0.9859	0.9915	0.9916	10.25	0.00014	0.00011
	93.46	96.62	6.54	3.38	93.9431	0.9662	0.9366	0.9346	0.9512	0.9513	10.29	0.00014	0.00011
	80.02	81.72	19.98	18.28	80.4286	0.8172	0.8035	0.8002	0.8103	0.8103	9.54	0.00013	0.00012
	86.56	84.48	13.44	15.52	86.9824	0.8448	0.8627	0.8656	0.8537	0.8537	9.57	0.00013	0.00012
	87.5	85.62	12.5	14.38	87.9281	0.8562	0.8726	0.875	0.8643	0.8644	9.58	0.00013	0.00012
	88.68	86.53	11.32	13.47	89.1127	0.8653	0.8843	0.8868	0.8747	0.8748	9.62	0.00013	0.00012
	91.34	89.02	8.66	10.98	91.7851	0.8902	0.9113	0.9134	0.9006	0.9007	9.65	0.00013	0.00012
	91.73	90.6	8.27	9.4	92.183	0.906	0.9164	0.9173	0.9111	0.9112	9.55	0.00013	0.00012
	96.05	94.33	3.95	5.67	96.5217	0.9433	0.9598	0.9605	0.9515	0.9515	9.61	0.00013	0.00012
	99.04	98.62	0.96	1.38	99.5331	0.9862	0.9904	0.9904	0.9883	0.9883	9.73	0.00013	0.00012
	95.27	99.84	4.73	0.16	95.7692	0.9984	0.9548	0.9527	0.9761	0.9763	9.78	0.00013	0.00012

**Table 8** Comparison of reliability, time complexity and error rate criteria

Correlation coefficient				
		R	T <sub>c</sub>	ER
<b>R</b>	Pearson correlation	1	-.239*	-.260**
	Sig. (two-tailed)		.013	.007
	N	108	108	108
<b>T<sub>c</sub></b>	Pearson correlation	-.239*	1	.862**
	Sig. (two-tailed)	.013		.000
	N	108	108	108
<b>ER</b>	Pearson correlation	-.260**	.862**	1
	Sig. (two-tailed)	.007	.000	
	N	108	108	108

\*Correlation is significant at the 0.05 level (two-tailed)  
 \*\*Correlation is significant at the 0.01 level (two-tailed)

different values for this colour space. In addition, new values have been obtained for these criteria according to the other colour spaces that will be implemented. Eventually, 108 algorithms were obtained on the basis of the 14 colour spaces that were adapted. However, the final number of the colour spaces that appeared in the decision matrix is 12. The reason is that the colour space groups of HSI, HSL and HSV, where luminance element is deleted from each colour space (I, L and V) and the chroma element (HS), is retained, which is common between colour space groups. Thus, the processing is applied only once. Table 7 illustrates the evaluation result for 108 colour space samples tested 13 of the criteria.

The relationship between all criteria will be determined according to statistical methods using the Pearson formula, as well as conducting the performance analysis for each

**Table 9** Comparison amongst sub-criteria as parameter matrix, relationship and behaviour

Correlation coefficient				
		MP	RP	BP
<b>MP</b>	Pearson correlation	1	-.973**	-.953**
	Sig. (two-tailed)		.000	.000
	N	108	108	108
<b>RP</b>	Pearson correlation	-.973**	1	.997**
	Sig. (two-tailed)	.000		.000
	N	108	108	108
<b>BP</b>	Pearson correlation	-.953**	.997**	1
	Sig. (two-tailed)	.000	.000	
	N	108	108	108

\*\*Correlation is significant at the 0.01 level (two-tailed)

criterion to determine their behaviour based on various threshold values that will be discussed in detail in the next sections.

**3.1.2 Correlation between criteria results**

The details of the correlation tests conducted amongst these criteria are presented in the next subsections.

**3.1.2.1 Correlation analysis in layer 1** This layer includes three independent criteria, namely reliability criterion, time complexity criterion and error rate criterion. The Pearson method is implemented to determine the extent of the relationship and correlation degree between the criteria. After conducting the test and selecting the desired path to determine the correlation between the criteria, we obtain the following results, as shown in Table 8.

**Table 10** Comparison training and validation sub-criteria

Correlation coefficient		T	V
T	Pearson correlation	1	.942**
	Sig. (two-tailed)		.000
	<i>N</i>	108	108
V	Pearson correlation	.942**	1
	Sig. (two-tailed)	.000	
	<i>N</i>	108	108

\*\*Correlation is significant at the 0.01 level (two-tailed)

Table 8 illustrates the relationship and degree of correlation between the criteria based on the rules of the correlation coefficient according to the Pearson test. Therefore, the correlation coefficient between the reliability and time complexity group was  $-0.239$ , indicating that a reverse correlation exists when ( $r < 0$ ) at a degree of significance of 0.013 for 108 samples. In addition, a high correlation exists in the negative aspect whilst the correlation value in the error rate group was  $-0.260$ , in which a reverse correlation exists when ( $r < 0$ ) at a degree of significance of 0.007 for 108 samples. Furthermore, the correlation degree between the time complexity and error rate groups was 0.892, indicating that a positive correlation exists when ( $r > 0$ ) at a degree of significance of 0.000 for 108 samples. Thus, a high correlation exists in the positive aspect. Overall, the existing correlation between each criterion is proven on the basis of the rules of Pearson test results.

**3.1.2.2 Correlation analysis in layer 2** The second layer includes two sub-criteria groups. The first group includes

three basic sub-criteria generated by the reliability group. The second group includes two sub-criteria derived from the error rate group. Pearson test will be used to calculate the correlation coefficient between each group (Table 9).

Table 9 highlights the result of correlation between three sub-criteria, namely the matrix, relationship and behaviour of parameters, which are all derived from the reliability group.

Table 10 highlights the correlation between validation and training sub-criteria, which are derived from the error rate within the dataset group. Overall, the existing correlation between each criterion in Tables 9 and 10 is proven on the basis of the Pearson test results.

**3.1.2.3 Correlation analysis in layer 3** Layer 3 comprises three groups of sub-sub-criteria. The first group includes four parameters generated from the parameter matrix, namely TP, FP, TN and FN. The second group includes four parameters derived from the parameter relationship, namely accuracy, precision, recall and specificity. The third group includes two parameters, namely *F*-measure and *G*-measure, which are generated from the parameter behaviour. Table 11 shows that the Pearson test is implemented to calculate the correlation coefficient for each group.

Table 11 highlights the correlation between four sub-sub-criteria, namely TN, FP, TN and FN, which are derived from the parameter matrix.

Table 12 highlights the correlation between four sub-sub-criteria, namely accuracy, recall, precision and specificity, which are derived from the relationship of parameters.

Table 13 highlights the correlation between two sub-sub-criteria, namely the *F*-measure and the *G*-measure,

**Table 11** Comparison amongst TP, FP, TN and FN sub-sub-criteria

Correlation coefficient		TP	FP	TN	FN
TP	Pearson correlation	1	-.744**	.744**	-1.000**
	Sig. (two-tailed)		.000	.000	.000
	<i>N</i>	108	108	108	108
FP	Pearson correlation	-.744**	1	-1.000**	.744**
	Sig. (two-tailed)	.000		.000	.000
	<i>N</i>	108	108	108	108
TN	Pearson correlation	.744**	-1.000**	1	-.744**
	Sig. (two-tailed)	.000	.000		.000
	<i>N</i>	108	108	108	108
FN	Pearson correlation	-1.000**	.744**	-.744**	1
	Sig. (two-tailed)	.000	.000	.000	
	<i>N</i>	108	108	108	108

\*\*Correlation is significant at the 0.01 level (two-tailed)

**Table 12** Comparison of accuracy, precision, recall and specificity sub–sub-criteria

Correlations		ACC	PR	RE	SP
ACC	Pearson correlation	1	1.000**	.741**	1.000**
	Sig. (two-tailed)		.000	.000	.000
	<i>N</i>	108	108	108	108
PR	Pearson correlation	1.000**	1	.744**	1.000**
	Sig. (two-tailed)		.000	.000	.000
	<i>N</i>	108	108	108	108
RE	Pearson correlation	.741**	.744**	1	.744**
	Sig. (two-tailed)		.000	.000	.000
	<i>N</i>	108	108	108	108
SP	Pearson correlation	1.000**	1.000**	.744**	1
	Sig. (two-tailed)		.000	.000	.000
	<i>N</i>	108	108	108	108

\*\*Correlation is significant at the 0.01 level (two-tailed)

**Table 13** Comparison between *F*-measure and *G*-measure sub–sub-criteria

Correlation coefficient		<i>F</i>	<i>G</i>
<i>F</i>	Pearson correlation	1	.951**
	Sig. (two-tailed)		.000
	<i>N</i>	108	108
<i>G</i>	Pearson correlation	.951**	1
	Sig. (two-tailed)		.000
	<i>N</i>	108	108

\*\*Correlation is significant at the 0.01 level (two-tailed)

which are derived from the behaviour of parameters. Overall, the existing correlation between each sub–sub-criterion in Tables 11, 12 and 13 is proven by the Pearson test results.

Overall, the Pearson test shows a strong correlation between the criteria based on the results. According to the results in Tables 8, 9, 10, 11, 12 and 13, the results demonstrate statistically significant differences between the criteria within each layer, emphasising the need to use all the criteria and sub-criteria in the proposed decision matrix that supported the proposed methodology.

### 3.1.3 Performance analysis of criteria results

Details of the performance analysis are presented below:

**3.1.3.1 Reliability group** In this step, we highlight the performance of reliability group based on nine threshold

values for each colour space used. Reliability group includes three key sections: (1) the matrix of parameters [true negative (TN), true positive (TP), false positive (FP) and false negative (FN)], (2) the relationship of parameters (accuracy, recall, precision and specificity) and (3) the behaviour of parameters (*F*-measure and *G*-measure).

#### A. Matrix of parameters

The matrix of parameters, which is the first and most important section in the reliability group, includes four key parameters, namely TN, TP, FP and FN. The matrix of parameters is also one of the evaluation techniques for skin detection approaches [69].

##### A.1 True negative criterion

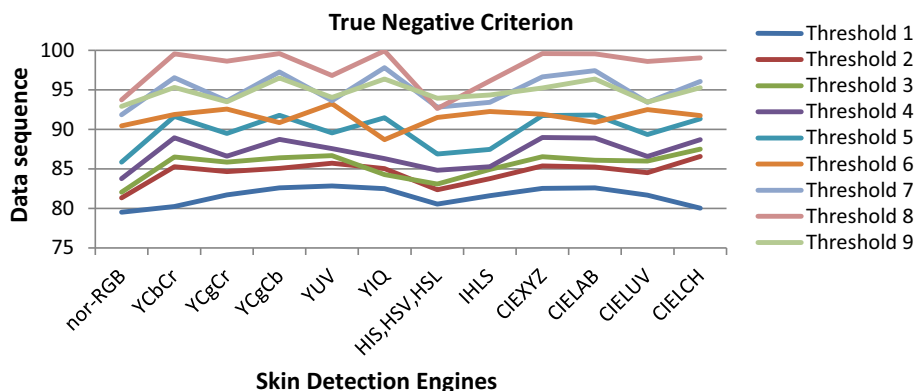
In this section, the behaviour of the first criterion within the matrix of parameters is discussed and analysed. Figure 19 illustrates the behaviour of TN criterion based on nine threshold values with different colour spaces.

Figure 19 illustrates the behaviour of the TN criterion at different threshold values with various colour spaces used as alternatives in the study. The graph shows the behaviour of this criterion to appear similar at each threshold value. However, the figure shows the lowest threshold value of the criterion at 79.5%, whereas the highest value is at 99.04%. The path of each threshold is nearly similar according to the colour spaces, except for thresholds 1 and 6. The behaviour of the criterion is shown at the value of threshold 1, which starts to slightly increase from Norm-RGB to YIQ. The value slightly drops at the HIS, HSV and HSL, slightly increases until CIELUV and drops again in CIELCH. By contrast, thresholds 2 and 3 exhibit the same behaviour, which starts rising from Norm-RGB to YCbCr and then stabilises their track to HIS, HSV and HSL. Such behaviour begins to slightly decline, slightly increases until CIEXYZ and gradually increases its track even CIELCH. Thresholds 4, 5, 7 and 9 start to slightly increase from Norm-RGB to YCbCr and begin to drop and increase as well as CIELCH. Meanwhile, threshold 6 has a reverse behaviour wherein the threshold starts to slightly increase from Norm-RGB to YCgCr, starts to fall and then increases to YIQ. The threshold slightly increases and then stabilises its track until CIELUV and then falls until CIELCH. Finally, threshold 8 starts to slightly increase from Norm-RGB to YCbCr and then begins to increase, thereby dramatically falling and sharply dropping in HIS, HSV and HSL, and starts to slightly increase in CIEXYZ where its track settled down to the end.

##### A.2 True positive criterion

Figure 20 shows the behaviour of TP criterion according to the changes in the threshold values and colour space.

**Fig. 19** Behaviour of true negative criterion with different colours



The behaviour is different from the previous standard as shown below.

Figure 20 illustrates the behaviour of the TP criterion at different threshold values with various colour spaces used as alternatives in this research. The figure shows the lowest threshold value of the criterion at 80.06%, whereas the highest value is 99.84%. Nearly all threshold values start at one point for Norm-RGB. Thus, the behaviour of this criterion is affected according to the changes in the threshold track. Threshold 1 starts to slightly increase until YCbCr, and its track evenly stabilises to YIQ, slightly increases until IHLS and stabilizes until the end. Thresholds 2 and 3 begin to slightly increase until YCbCr, stabilise their track until YIQ, slightly drop at HIS, HSV and HSL and then slightly increase until the end of its track. Thresholds 4, 5 and 7 have nearly similar tracks from start to end. These thresholds start to increase to YCbCr, where tracks change between high and low until YIQ and then drop to a minimum value of HIS, HSV and HSL. The thresholds increase again until CIEXYZ; thus, their track stabilises until CIELAB, slightly increase and decrease to the end of their tracks. Threshold 9 has a similar track with the previous threshold but sharply drops at HIS, HSV and HSL; the threshold sharply increases to CIEXYZ, settles in CIELAB and then increases to its end. However, threshold 6 has a different track, which evenly increases to YCbCr, settles in

YCgCr and changes its track between high and low until YIQ. The track then drops to the lowest value at HIS, HSV and HSL, increases to CIELAB and slightly decreases until the end. Finally, threshold 8 represents the highest threshold value, which begins to increase until YCbCr, stabilises until it sharply drops to the lowest value at HIS, HSV and HSL, and increases again until CIEXYZ; its track stabilises until the end.

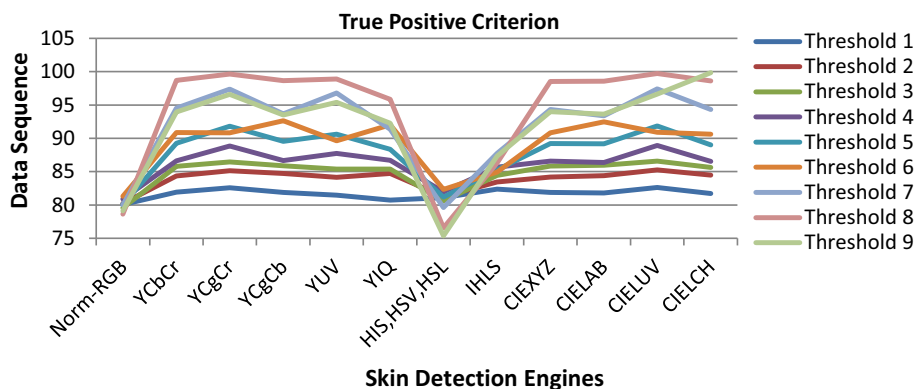
**A.3 False positive criterion**

Figure 21 shows the behaviour of FP criterion based on different threshold values with various colour spaces as shown in the graph below.

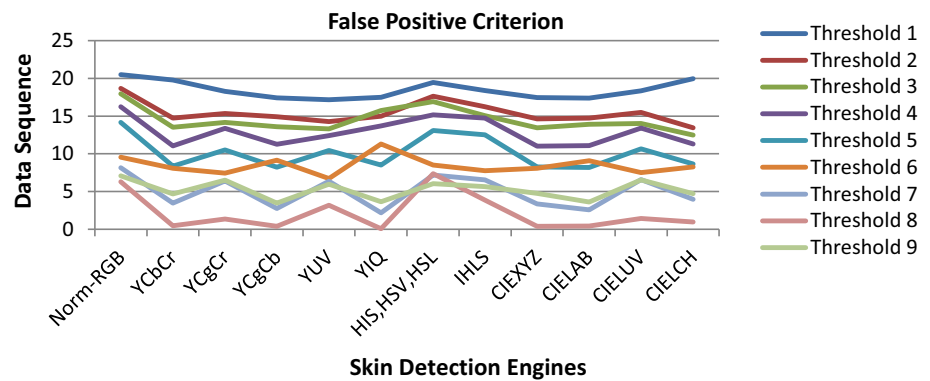
Initially, this criterion is complementary to the true negative criterion within probabilistic parameters of the reliability group. However, the figure shows that the lowest threshold value of the criterion is 0.07%, whereas the highest value is 20.5%. Thus, the track of each threshold is nearly similar according to the colour spaces, except for thresholds 1 and 6.

Threshold 1 has the highest value, which slightly declines until YIQ and then slightly increases even in HIS, HSV and HSL, gradually dropping until CIELAB and then slightly increasing again to the end. Thresholds 2 and 3 have a similar track where they slightly decline even in YCbCr and then continue to increase until YUV; the

**Fig. 20** Behaviour of the true positive criterion with different colour spaces



**Fig. 21** Behaviour of the false positive criterion with different colour spaces



thresholds slightly increase even in HIS, HSV and HSL, slightly drop again and then straighten their track until the CIELUV where they drop to the end. In addition, thresholds 4 and 5 have the same track, where they start to slightly drop even in YCbCr and then fluctuate until YIQ. These tracks slightly increase even in HIS, HSV and HSL and slightly decline. Finally, their tracks settle down to CIEXTZ and then slightly increase and decrease again until the end. Threshold 6 has a different behaviour, in which its track drops until YCgCr and gradually increases and decreases, slightly increases in YIQ and slightly drops until it stabilises to the end. By contrast, thresholds 7 and 9 have tracks similar to that of threshold 5. Finally, threshold 8 records the lowest threshold value, where its track slightly drops at YCbCr and then gradually increases and decreases until YIQ. Its track sharply increases and then drops even in CIEXYZ until its track remains stable to the end. The general characteristic of this criterion is that its behaviour is similar to the TN criterion behaviour but in the opposite direction.

#### A.4 False negative criterion

Figure 22 illustrates the behaviour of the FN criterion using nine thresholds with various colour spaces as shown in the chart below.

Initially, this criterion is complementary to the TP criterion within probabilistic parameters of the reliability group. However, the lowest threshold value of the criterion is 0.16%, whereas the highest value is 24.63%. Figure 22 shows that nearly all threshold values start at one point for Norm-RGB. Thus, the behaviour of this criterion is affected according to the changes in the threshold track. Threshold 1 slightly declines until YCgCr and slightly increases to stabilise even at YIQ and drops to IHLS. The threshold slightly rises and declines at CIELUV and rises again to the end. Thresholds 2 and 3 nearly exhibit a similar trend as threshold 1, but they sharply rise at HIS, HSV and HSL, gradually drop until CIELUV and slightly rise to the end afterwards. Thresholds 4 and 5 start similarly as the previous thresholds, in which they sharply rise

even to HIS, HSV and HSL, gradually decline until CIELUV and increase again to the end. Threshold 6 exhibits a behaviour that is different from others when it starts from the same point and gradually declines until YcbCr. This threshold moderates its track at YcgCr, slightly drops to YcgCb, rises and declines again. Finally, it sharply rises to the top even at HIS, HSV and HSL afterwards, gradually dropping to the end. Thresholds 7 and 9 exhibit similar behaviour where they start to decline from the starting point until YCgCr, change their track up and down and rise to the top even at HIS, HSV and HSL. Threshold 7 gradually drops to the end whilst threshold 9 continues to decline. Finally, threshold 8 records the lowest threshold value, where it starts at the same point and continues to decline even at YCbCr and then stabilises its track to YUV. This threshold sharply rises to the top until HIS, HSV and HSL, sharply declines until CIEXYZ and gradually continues to the end. We conclude that the behaviour of this criterion is similar to that of TP criterion but in the opposite direction.

#### B. Relationship of parameters

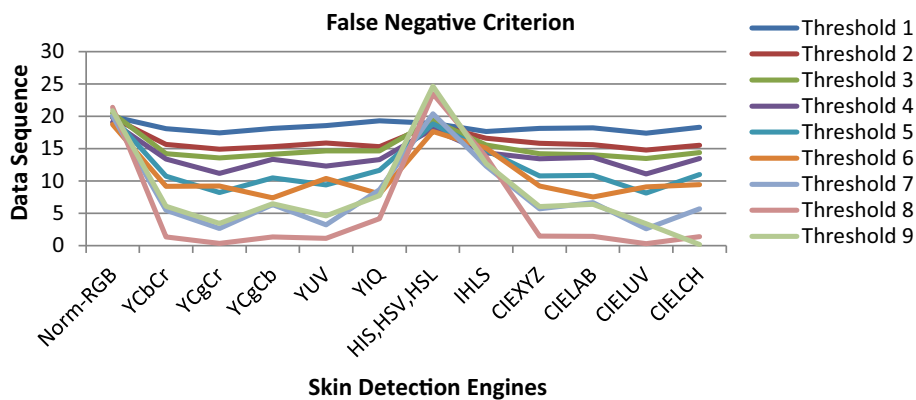
In this step, the second group of sub-criteria within the reliability group, namely relationship of parameters which includes (accuracy, recall, precision and specificity), is discussed and analysed.

##### B.1 Accuracy criterion

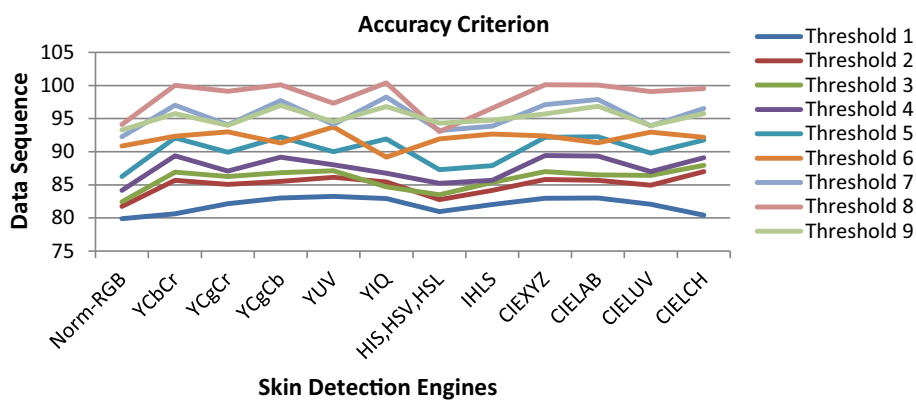
Accuracy refers to the exactness of the analytical method in correctly identifying outliers and inliers from total data [66].

Figure 23 shows that the criterion behaves similarly with that of TP criterion due to the convergence of the values between the two criteria. The accuracy criterion is an important measure of the relationship of parameters of the reliability group. However, the figure shows the lowest threshold value of the criterion at 79.90%, whereas the highest threshold value is 100.40%. Thus, the track of each threshold is nearly similar according to the colour spaces, except for thresholds 1 and 6.

**Fig. 22** Behaviour of the false negative criterion with different colour spaces



**Fig. 23** Behaviour of the accuracy criterion with different colour spaces



Threshold 1 records the lowest value, where it starts to slightly rise from Norm-RGB to YIQ and slightly drops at the HIS, HSV and HSL afterwards. It starts to slightly rise until CIELUV and drops again in CIELCH. Thresholds 2 and 3 exhibit the same behaviour, where they start to rise from Norm-RGB to YCbCr and then stabilise their track to HIS, HSV and HSL. These thresholds begin to slightly decline, gradually rise until CIEXYZ and gradually rise their tracks even to CIELCH. Thresholds 4, 5, 7 and 9 start slightly high from Norm-RGB to YCbCr and begin to drop and rise even at CIELCH. Threshold 6 demonstrates a reverse behaviour, where it starts to slightly rise from Norm-RGB to YCgCr, gradually falls, rises to YIQ, slightly rises and stabilises its track until CIELUV, thereby falling until CIELCH afterwards. Finally, threshold 8 starts at the highest value, slightly rises from Norm-RGB to YCbCr, begins to rise and dramatically falls, thereby sharply dropping at HIS, HSV and HSL. This threshold starts to slightly rise to CIEXYZ, and its track settles down to the end.

**B.2 Recall criterion**

The recall is an important criterion to measure the wholeness or quantity values. This criterion refers to the TP rate for the probability of complete retrieval of the parameter values. [71]. Figure 24 shows all threshold

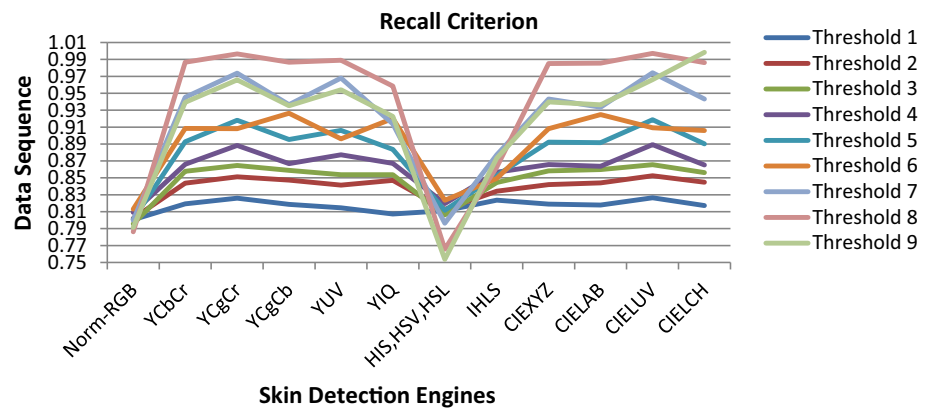
values starting at approximately the same point from Norm-RGB. This criterion behaves fairly similar to that of the TP criterion due to the matching in the track of thresholds, as shown in the graph. The reason behind that recall is the ratio of TP components to elements inherently ranked as positive. Thus, recall represents the number of correctly classified positive examples divided by the number of positive examples in the data. Finally, the contribution of the recall represents almost similar to the TP.

Recall is a measure of completeness or quantity. It is the average probability of a complete retrieval referred to as the TPR. Recall is the ratio of TP components to elements inherently ranked as positive. Thus, recall represents the number of correctly classified positive examples divided by the number of positive examples in the data. Finally, the contribution of the recall also focuses only on the positive examples and predictions.

Figure 24 shows that the lowest threshold value of the recall criterion is 0.791%, whereas the highest threshold value is 0.997%. This figure illustrates the behaviour of the recall criterion at different threshold values with various colour spaces used as an alternative in the study. The chart demonstrates that nearly all threshold values start at one point from Norm-RGB. Thus, the behaviour of this criterion is affected according to the changes in the



**Fig. 24** Behaviour of the recall criterion with different colour spaces



threshold track. Threshold 1 starts to slightly increase until YCbCr; its track stabilises even in YIQ and then slightly increases until IHLS; finally, its track stabilises to the end. Meanwhile, thresholds 2 and 3 begin to slightly increase until YCbCr and then stabilise their track until YIQ. These thresholds slightly drop at HIS, HSV and HSL and then slightly increase to the end. Thresholds 4, 5 and 7 nearly have a similar track from start to end. These thresholds start to increase to YCbCr. Their tracks change between up and down until YIQ and then drop to a minimum value of HIS, HSV and HSL. Finally, their tracks increase again until CIEXYZ and stabilise until CIELAB, slightly increase and decrease to the end. Threshold 9 has a track similar to that of the previous threshold. However, its track sharply drops at HIS, HSV and HSL and sharply increases to the CIEXYZ. The track remains stable until CIELAB and reaches its maximum height to the end. Threshold 6 has a different track, where it increases even in YCbCr and settles up until YCgCr. Its track changes between up and down until the YIQ and drops to the lowest value at HIS, HSV and HSL. The track of this threshold increases to CIELAB and then slightly decreases until the end. Finally, threshold 8 records the highest threshold value, which begins to increase until YCbCr and stabilises until it sharply drops to the lowest value at HIS, HSV and HSL. Its value increases again until CIEXYZ, and its track stabilises until the end.

### B.3 Precision criterion

Precision is also an important measure and refers to the ratio of information relevant to all information gathered through networks or different sensors and services [72].

Figure 25 depicts that precision has a behaviour similar to accuracy due to the convergence of the respective values. The precision criterion is an important measure of the relationship of parameters of the reliability group. However, the figure shows that the lowest threshold value of precision is 0.796%, whereas the highest value of the threshold is 0.996%. Thus, the track of each threshold is

nearly similar according to the colour spaces, except for thresholds 1 and 6. Threshold 1 records the lowest value, which slightly starts to increase from Norm-RGB to YIQ. The value slightly drops at the HIS, HSV and HSL, slightly increases until CIELUV and decreases again in CIELCH. By contrast, thresholds 2 and 3 have the same behaviour, in which their values start to increase from Norm-RGB to YCbCr and then stabilise their track to HIS, HSV and HSL. Their tracks begin to slightly decline, slightly increase until CIEXYZ and gradually increase even in CIELCH. Thresholds 4, 5, 7 and 9 start to slightly increase from Norm-RGB to YCbCr and then begin to drop and increase even in CIELCH. Threshold 6 has a reverse behaviour, in which it starts to slightly increase from Norm-RGB to YCgCr and starts to decrease and increase to YIQ. This threshold slightly increases and stabilises its track until CIELUV and decreases until CIELCH. Finally, threshold 8 starts at the highest value, which slightly increases from Norm-RGB to YCbCr, increases and dramatically and sharply drops at HIS, HSV and HSL. This threshold starts to slightly increase to the CIEXYZ until its track settles down to the end.

### B.4 Specificity criterion

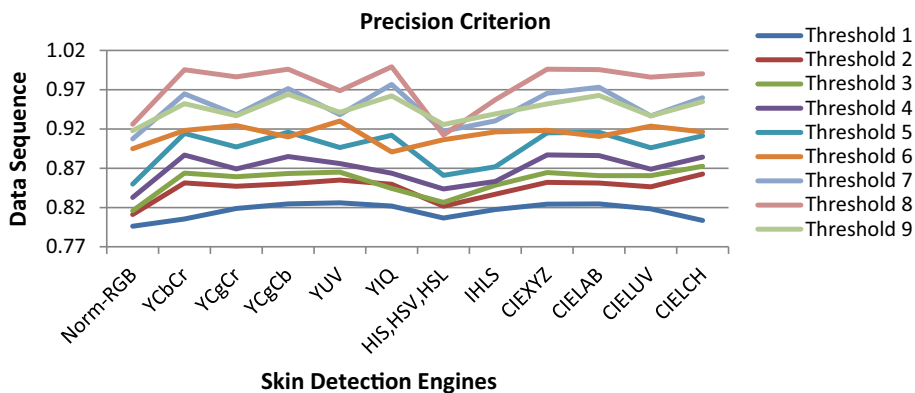
Specificity is an important measure of the relationship of parameters that represents the capability of a classifier to distinguish between classes of the negative value from 0 to 1 [70]. The specificity criterion behaviour is shown in Fig. 26.

Figure 26 illustrates that the behaviour of specificity is similar to that of precision at the threshold values according to different colour spaces.

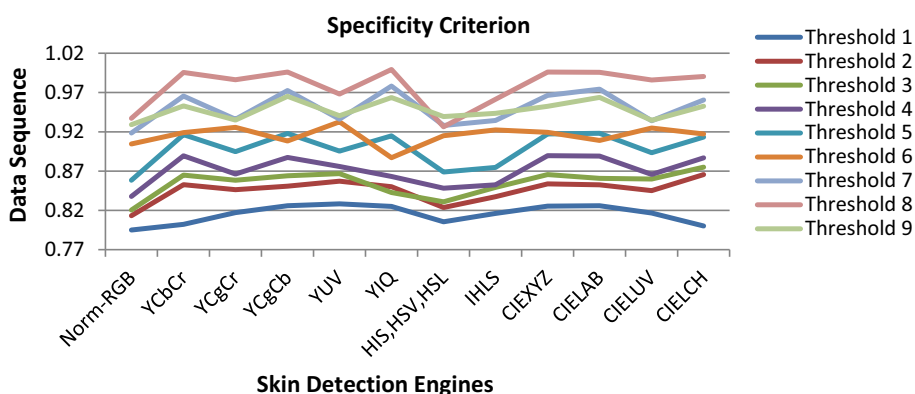
### C. Behaviour of parameters

The last part of the reliability group represents the behaviour of parameters and includes two key parameters, namely *F*-measure and *G*-measure. This group measures and tests the behaviour of the parameters that are closely related to those of the precision and recall.

**Fig. 25** Behaviour of the precision criterion with different colour spaces



**Fig. 26** Behaviour of the specificity criterion with different colour spaces



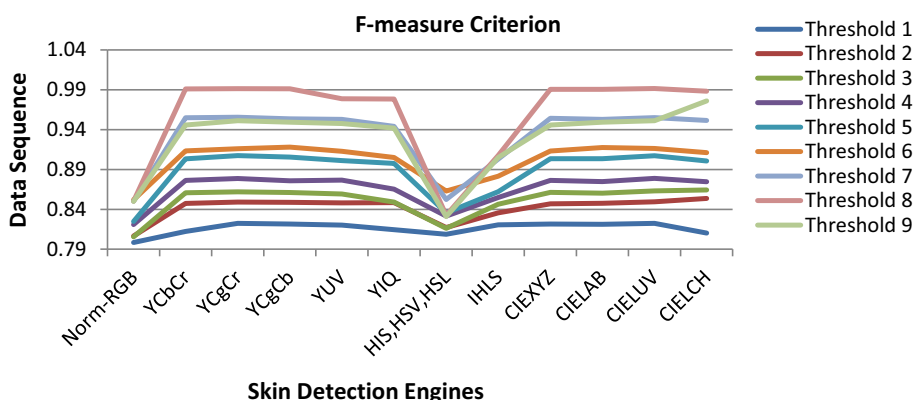
C.1 *F-measure criterion*

*F-measure* is a simple method to represent the precision and recall, which is considered as the most popular criterion for evaluating classification quality [73].

Figure 27 shows the behaviour of *F-measure* criterion based on nine threshold values with different colour spaces as shown below. The figure shows that the lowest threshold value of the criterion is at 0.798%, whereas the highest threshold value is at 0.991%. Threshold 1 starts to slightly increase at the lowest value from Norm-RGB until YCbCr and then stabilizes its track until the figure shows the lowest threshold value of the criterion at 0.796%.

Meanwhile, the highest threshold value is at 0.996%, which slightly increases and stabilises until CIELUV and then slightly drops until the end track. Thresholds 2 and 3 start at the same point and slightly increase up to YCbCr. Their track stabilises until YIQ and descends at HIS, HSV and HSL, thereby slightly increasing up to CIEXYZ and then stabilises their track again until the end. Thresholds 4 and 5 begin to slightly increase from YCbCr and then stabilize their path until YIQ. These thresholds decline their tracks at HIS, HSV and HSL and then increase until CIELUV, slightly dropping to the end. Finally, thresholds 6, 7, 8 and 9 begin to increase from the same point where threshold 6

**Fig. 27** Behaviour of the *F-measure* criterion with different colour spaces



starts to increase until YCbCr and then settles down to YIQ. The thresholds sharply fall at HIS, HSV and HSL, increase to CIEXYZ and settle down their track until the end. Thresholds 7 and 9 have the same track as that of the previous threshold. They sharply decrease at HIS, HSV and HSL and sharply increase at CIEXYZ. The track of threshold 9 stabilises to the end whilst threshold 9 increases from CIELUV. Finally, threshold 8 records the highest value, which starts to increase to YCbCr and stabilises its track until YCGCb. The value slightly drops at YUV and continues to sharply drop at HIS, HSV and HSL. Finally, its value sharply increases at CIEXYZ and stabilises its track to the end.

C.2 G-measure criterion

The second criterion within the behaviour of parameters defined as G-measure, which refers to the geometric mean for values of precision and recall or the general measure for classification of algorithms based on the performance and accuracy of the measured sample classification.

Figure 28 shows that G-measure exhibits a behaviour similar to that of F-measure due to the symmetry in their final values, which affect the behaviour of G-measure according to the threshold distribution values at each colour space used.

3.1.3.2 Time complexity criterion Time complexity criterion is the time processing algorithm for all image samples needed during image segmentation process based on the size of the image [74]. In this section, the key criterion of time complexity is discussed according to the distribution of threshold values with different colour spaces.

Figure 29 shows that the criterion has the lowest threshold value at 8.01%, whereas the highest threshold value is at 12.42%, which indicates that the threshold values for this criterion are nearly identical. Consequently, the matching in its values clearly influences the behaviour of this criterion. Thus, the track of threshold values appears to be identical from the starting point to the end of its track.

The threshold values start at the lowest value, which sharply increases from Norm-RGB until YCgCr and then continues to slightly increase until YIQ. The values sharply drop until HIS, HSV and HSL. Finally, they start to slightly increase until CIEXYZ, and their tracks stabilise until CIELAB and slightly decrease until the end of its track.

3.1.3.3 Error rate within dataset The error rate is the minimum of error rate measured during the implementation of the training process based on the dataset used by an irreducible classifier which implemented the training and validation process [69].

A. Error rate of validation criterion

A cross-validation process, which is widely used in this study, is used to set the error rate for the training data. The dataset is divided into three sections, namely the majority of data for training and validation and testing.

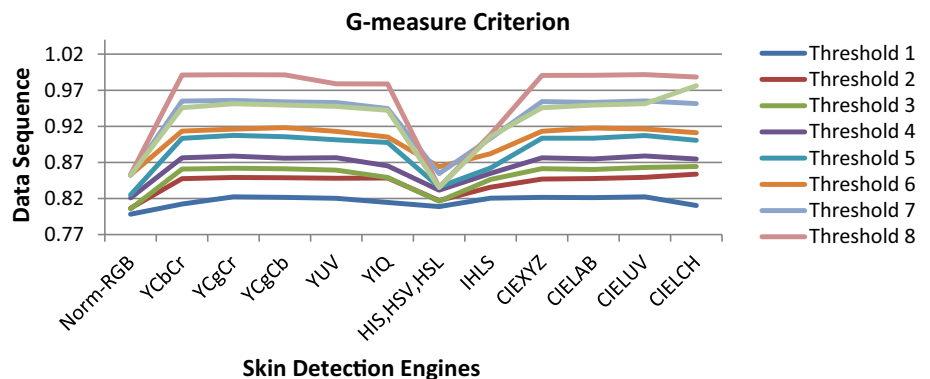
Figure 30 illustrates the measurement of the error rate for validation criterion based on nine thresholds according to different colour spaces. The threshold values are identical for each colour space. Thus, the track of the threshold values starts at the lowest value from Norm-RGB and sharply increases until YCgCr. It slightly increases to YIQ and sharply declines until HIS, HSV and HSL. It sharply increases to CIEXYZ. It slightly rises until CIELAB and slowly drops to the end of the track.

B. Error rate of training criterion

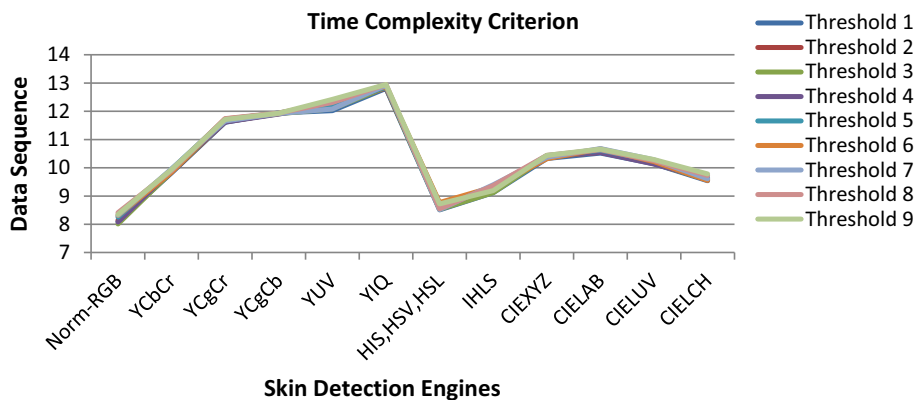
The training set is also an important stage in calculating the error rate during the dividing process of the dataset. Figure 31 shows the behaviour of the training criterion.

Figure 31 shows the behaviour of the error rate of training criterion according to the distribution of threshold values at each colour space. This figure exhibits a status similar to that of the validation criterion through matching threshold values according to each colour space. The values of the threshold start from Norm-RGB, sharply rise until YCgCb and slightly drop until YUV. These values

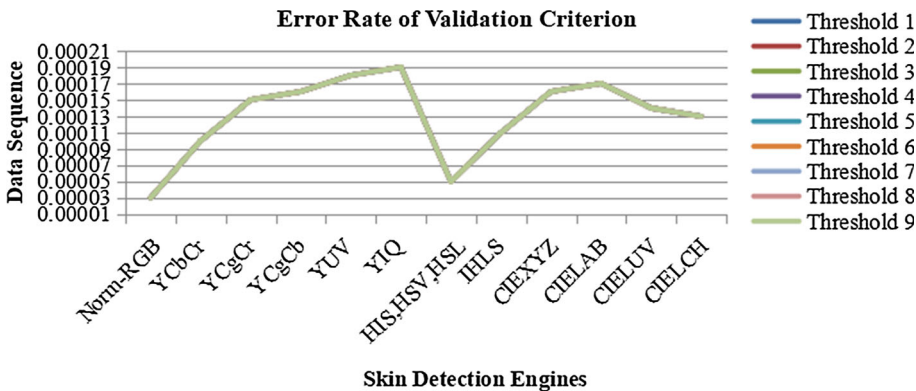
Fig. 28 Behaviour of the G-measure criterion with different colour spaces



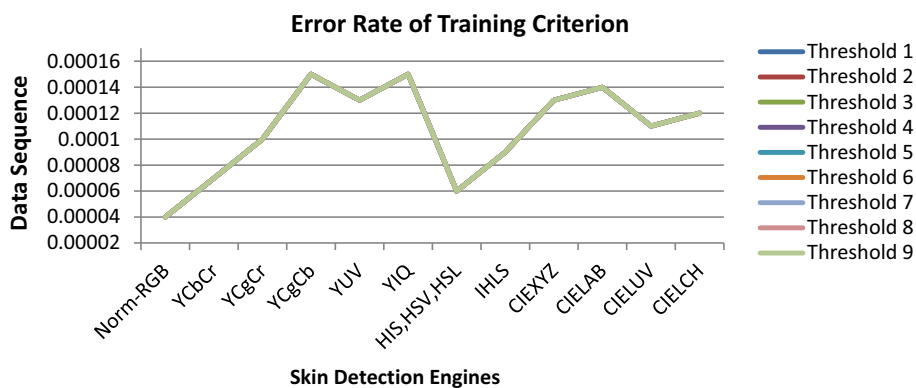
**Fig. 29** Behaviour of the time complexity criterion with different colour spaces



**Fig. 30** Behaviour of the error rate of validation criterion with different colour spaces



**Fig. 31** Behaviour of the error rate of training criterion with different colour spaces



slightly rise to YIQ, sharply decline to HIS, HSV and HSL and sharply rise again to CIEXYZ. They continue to slightly rise to the CIELAB, slightly decline to CIELUV and slightly rise until the end of their track.

In conclusion, the behaviour of criteria in all scenarios is affected by the distribution of threshold values for each criterion according to the different colour spaces used. The reliability group has three main sections. The first section represents a matrix of parameters, which includes probabilistic parameters that have nearly identical behaviour as in the charts. The TN criterion and FP criterion have a similar behaviour but in the opposite direction due to the values of the FP criterion, which are complementary to the

TN criterion. The TP criterion and FN criterion have a similar behaviour for the same reason. The second section represents a relationship of parameters including the four main criteria. The behaviour of the criteria is affected by the values of the matrix of parameters according to their measurement. The third section includes two criteria, namely *F*-measure and *G*-measure. Both criteria have similar behaviour due to the convergence of their values. By contrast, the time complexity has a specific behaviour as shown in the diagram. The behaviour of the time complexity is clearly affected by the large convergence between its values, which is distributed according to the threshold values compared with the different colour spaces.

Finally, the error rate within the dataset has two basic criteria, namely validation and training. The behaviour of the two criteria is clearly affected by the matching between the threshold values at each colour space for these criteria. Therefore, the results show differences between the behaviours of criteria, thereby emphasising the need to use all the criteria as multidimensional measurements in the methodology proposed.

### 3.2 Benchmarking of real-time IoT skin detectors

In this stage, the process of integration is implemented between the best MCDM techniques according to the proposed methodology. AHP is performed to generate different weights according to the criteria used in our study. This method is based on the preferences of six evaluators from different universities in Malaysia using pairwise question method to provide the results of the final weights. By contrast, the TOPSIS method is employed to generate the final results based on the obtained weights from the method and the performed decision matrix. The results are based on the two main contexts, namely individual and group contexts. Thus, the selection of a suitable context is recommended based on experiment implementation and different aggregation processes conducted to achieve the selection procedure. These contexts are emphasised based on individual decision-making for decision-makers and group decision-making for multiple decision-makers. The results are obtained through the integration process for selecting the best alternatives, which is considered the main objective of this section. Consequently, the results for the six evaluators are presented and discussed in detail.

#### 3.2.1 Multi-layer weight measurement using AHP

Table 14 presents the results based on the preferences of the six evaluators from the Malaysian universities. Questions are presented according to the rules of the hierarchical analysis process. ML-AHP is applied to generate standard weights according to the preferences of the evaluators. The first evaluator presents the percentage of the reliability group at 57.3%, the time complexity group percentage at 35.3, and the error rate group at 7.4%. The second evaluator presents the percentage of the reliability group at 28.1%, the time complexity group at approximately 8.1% and the error rate group at 63.8%. The third evaluator presents the percentage of the reliability group at 33.3%, the time complexity group at 33.3% and the error rate group at 33.4%. The fourth evaluator presents the percentage of the reliability group at 62.2%, the time complexity group at 30.2% and the error rate group at 7.6%. The fifth evaluator presents the percentage of the reliability

group at 23.9%, the time complexity group at 13.8% and the error rate group at 62.3%. The last evaluator presents the percentage of the reliability group at 21.1%, the time complexity group at 68.6% and the error rate group at 10.2%. The results of the criteria weights gathered from the six evaluators are utilised to complete the decision matrix that will be used in decision-making.

#### 3.2.2 TOPSIS performance based on different evaluators' weights

In this section, TOPSIS is used to evaluate and benchmark the skin detector approach on the basis of 108 colour space algorithms from the perspective of the six evaluators. Thus, this procedure is applied to select the best method. Table 14 provides the preference weights of the features with the appropriate colour space algorithm from the perspective of the evaluators. Six experts conduct pairwise comparisons to evaluate the degree of importance of each evaluation criterion. TOPSIS is implemented to distinguish the worst and best performances of the skin detection approach for each experiment. The results from these experiments are compared with the ideal and worst performances.  $S^-$  and  $S^+$  represent the approaches whose performances are closest to the worst and ideal performances, respectively. Based on the TOPSIS rules, the approach closest to the best performance and farthest from the worst performance is selected as the ideal approach. The preferences of the six evaluators are as follows.

According to the results of ML-AHP, the weights given by the first evaluator for the reliability group, represented in different parameters, are 41.5%, 12.8% and 3% for TP, FP, TN and FN (accuracy, recall, precision and specificity) and  $F$ -measure and  $G$ -measure, respectively. The time complexity group is given a weight of 35.3%, whereas the error rate group represented in training and validation is given weights of 6.2% and 1.2%, respectively. Each colour space algorithm is evaluated using different attributes. The results of the TOPSIS ranking indicate that the first evaluator gives an average of  $0.49763 \pm 0.13776$ . The value of the highest rank is 0.8080, whereas the lowest is 0.1527. Table 17 in Appendix presents the complete results from the first evaluator.

The weights given by the second evaluator for the reliability group represented in TP, FP, TN and FN (accuracy, recall, precision and specificity) and  $F$ -measure and  $G$ -measure are 2.5%, 8.5% and 17.1%, respectively. A weight of 8.1% is given for the time complexity group, whereas the error rate group is given the same weight of 31.9% for training and validation. The results from the second evaluator yield an average of  $0.38137 \pm 0.28451$ . The highest rank value is 0.955, whereas the lowest is

**Table 14** ML-AHP measurement for weight preferences

	TN	TP	FP	FN	Accuracy	Recall	Precision	Specificity	F-measure	G-measure	Time complexity	Validation	Training
UKM	0.104	0.104	0.104	0.104	0.032	0.032	0.032	0.032	0.015	0.015	0.354	0.012	0.062
IIUM	0.006	0.006	0.006	0.006	0.021	0.021	0.021	0.021	0.085	0.085	0.081	0.319	0.319
UNITEN	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.056	0.056	0.333	0.167	0.167
UPM-1	0.084	0.084	0.084	0.084	0.025	0.025	0.025	0.025	0.093	0.093	0.302	0.008	0.068
UPM-2	0.008	0.008	0.008	0.008	0.034	0.034	0.034	0.034	0.034	0.034	0.138	0.468	0.156
UPM-3	0.036	0.036	0.036	0.036	0.005	0.005	0.005	0.005	0.024	0.024	0.686	0.077	0.026

0.117. Table 18 in Appendix presents the complete results from the second evaluator.

For the reliability group, the third evaluator gives a weight of 11.1% for each parameter. The time complexity group is given a weight of 33.3%, whereas the error rate group represented in training and validation is given a weight of 16.7% for each parameter. The results from the third evaluator yield an average of  $0.42512 \pm 0.22686$ . The highest rank value is 0.8588, whereas the lowest is 0.0453. Table 19 in Appendix presents the complete results from the third evaluator.

The weights given by the fourth evaluator for the reliability group represented in TP, FP, TN and FN (accuracy, recall, precision and specificity) and *F*-measure and *G*-measure are 33.7%, 9.9% and 18.6%, respectively. The time complexity group is given a weight of 30.2%, whereas the error rate group represented in training and validation is given weights of 6.8% and 0.8%, respectively. The results from the fourth evaluator yield an average of  $0.49409 \pm 0.13362$ . The highest rank value is 0.8009, whereas the lowest is 0.1475. Table 20 in Appendix presents the complete results from the fourth evaluator.

The weights provided by the fifth evaluator for the reliability group represented in TP, FP, TN and FN (accuracy, recall, precision and specificity) and *F*-measure and *G*-measure are 3.3%, 13.7% and 6.8%, respectively. A weight of 13.8% for the time complexity group is given. Training and validation parameters representing the error rate group are given weights of 15.6% and 46.8%, respectively. The results from the fifth evaluator attain an average of  $0.37469 \pm 0.28847$ . The highest rank value is 0.9678, whereas the lowest is 0.0093. Table 21 in Appendix presents the complete results from the fifth evaluator.

The weights given by the sixth evaluator for the reliability group represented in TP, FP, TN and FN (accuracy, recall, precision and specificity) and *F*-measure and *G*-measure are 14.5%, 1.9% and 4.8%, respectively. The time complexity group is given a weight of 68.6%, whereas the error rate group represented by training and validation is given 2.6% and 7.7%, respectively. The results of the colour space algorithms achieve an average of  $0.49286 \pm 0.22406$ . The highest rank value is 0.8373, whereas the lowest is 0.0552. Table 22 in Appendix presents the complete results from the sixth evaluator.

Further details are discussed in Fig. 32, representing the preferences of the evaluators’ rankings based on internal and external aggregation values.

According to the tables generated by applying the TOPSIS method for selecting ideal alternatives, the final results of internal aggregation vary. Therefore, these results are evaluated by each evaluator based on the similarities of their values, as shown in Fig. 32. The results for the first

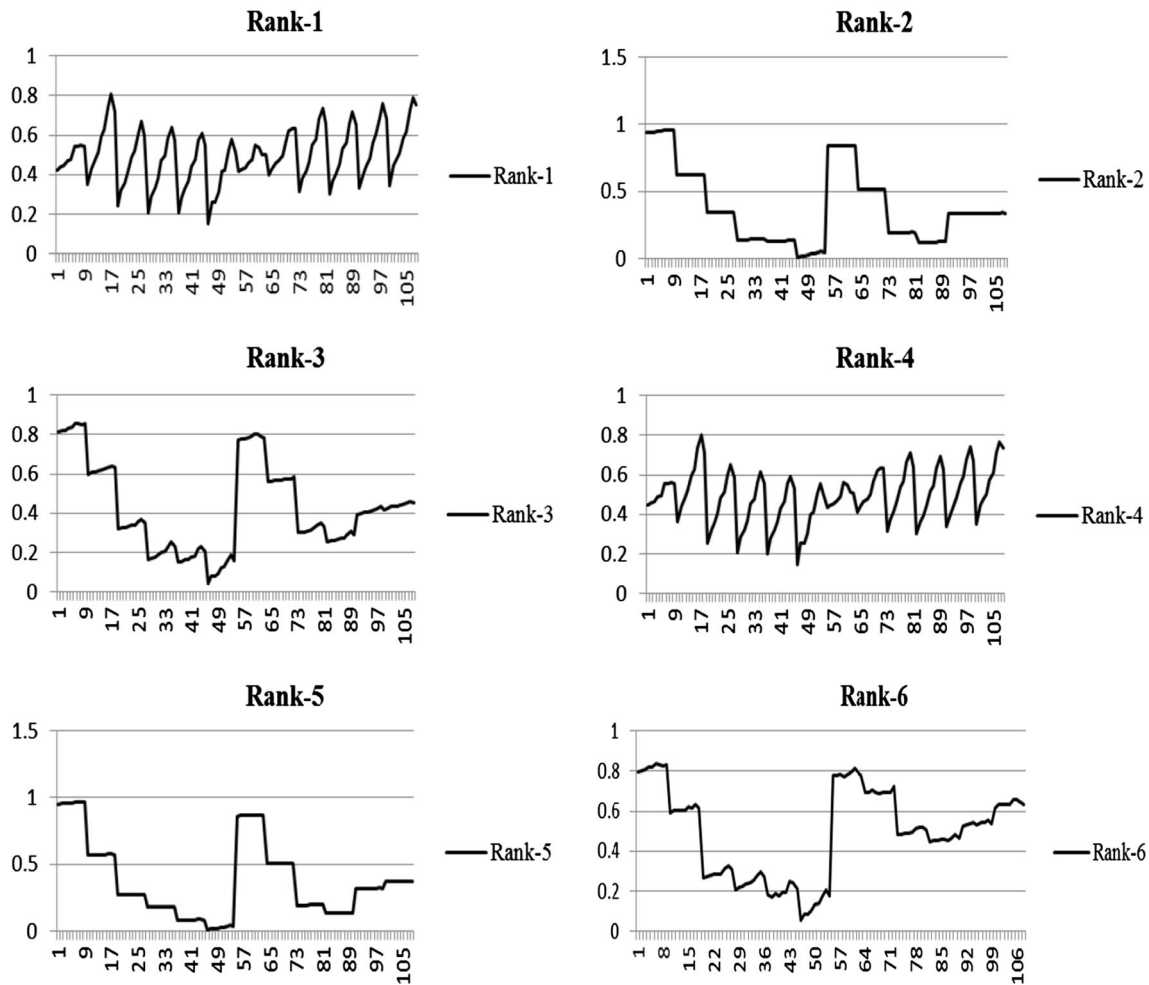


Fig. 32 Virtualised ranking for six evaluators

and fourth evaluators show a similar preference in the YCbCr colour space. The results for the second and fifth as well as the third and sixth evaluators yield similar results by selecting the normalised RGB colour space. Therefore, most of the evaluators' results confirm the selection of the normalised RGB colour space. The details in the next section show the results for the internal and external aggregations within the group decision.

In summary, the results from the evaluators indicate a lack of consensus for making a joint decision because of the difficulty and complexity of individual decision-making. Thus, the group decision-maker context discusses and compares individual decisions based on the results. The next section shows the results for the internal and external aggregations within the group decision-making.

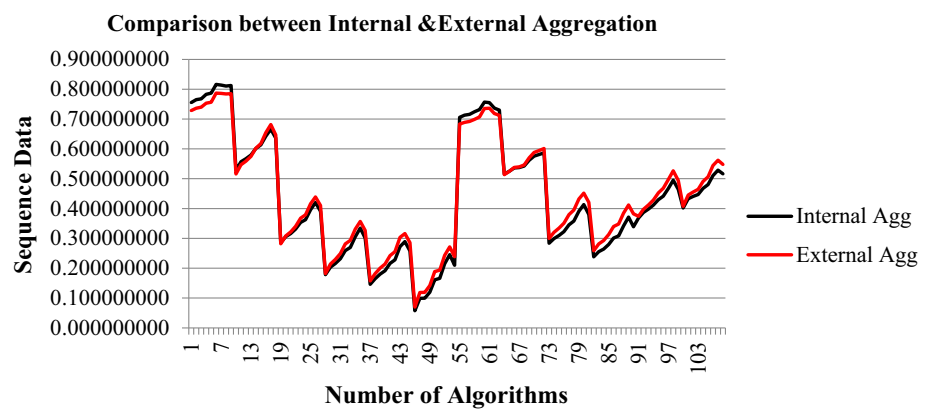
### 3.2.3 Group TOPSIS with internal and external aggregation

Many decision-making issues are resolved through cooperative efforts within organisations. However, according to

the two methods mentioned in the literature, TOPSIS expands to the group decision environment, either through internal or external aggregations. Internal aggregations aim to adopt the aggregation process in the separation phase. In such situations, group separation is conducted amongst various aggregating decision values based on the distance of the positive and negative ideals followed by the following process. Thus, the internal aggregation calculation based on the summation values of the negative separation is divided by the summation of negative separation values and the summation of positive separation values for all evaluators ( $S^- / (S^- + S^+)$ ), whereas external aggregation is determined by calculating the averages of all ranked values for all evaluators. Table 15 shows the results of group TOPSIS with the applied internal and external aggregations. The results of external aggregation show that high values are obtained in the normalised RGB colour space and demonstrate relatively identical internal and external aggregation values.

**Table 15** Group decision-maker of TOPSIS method with external and internal aggregations

Internal agg	External agg	Internal agg	External agg	Internal agg	External agg	Internal agg	External agg
0.75540413	0.728462921	0.20360009	0.214195117	0.715588808	0.691935664	0.279624839	0.312910362
0.765196872	0.736389302	0.21591285	0.229576432	0.724241963	0.698777958	0.300911904	0.339703176
0.767852619	0.739821909	0.23173244	0.249084501	0.731854128	0.707212241	0.308457646	0.348268337
0.782640989	0.753053356	0.25926195	0.281543083	0.756977802	0.735531853	0.341683829	0.384646775
0.78710423	0.756176802	0.26985588	0.293246805	0.754938985	0.73633807	0.371135821	0.412022354
0.816052934	0.786700332	0.30534859	0.330809909	0.736803272	0.718643343	0.338953272	0.382506261
0.813965619	0.785896345	0.33419751	0.356791553	0.729903493	0.712498735	0.368212545	0.37422124
0.810499041	0.784046586	0.30093977	0.326460009	0.515374436	0.513875895	0.386732112	0.397444093
0.812223751	0.785187243	0.14599179	0.157226071	0.524876249	0.524507387	0.398296954	0.41215902
0.528687669	0.515726373	0.16484707	0.181634526	0.53610646	0.537663155	0.411358516	0.428581196
0.556994391	0.546756829	0.18018585	0.200337559	0.537532174	0.539699114	0.429681672	0.452548606
0.567140466	0.559039624	0.19160121	0.213963775	0.542377533	0.545824282	0.441647021	0.467996274
0.579574405	0.574562659	0.21609028	0.242289732	0.561806889	0.570219026	0.467240628	0.49721201
0.601220114	0.602263027	0.22822913	0.255473124	0.575542102	0.588126927	0.495159404	0.526272567
0.613722011	0.617775486	0.27251061	0.303336236	0.581126593	0.594388763	0.464266919	0.494224097
0.642234098	0.654293665	0.28922547	0.316288837	0.586841973	0.60114385	0.402121301	0.407531084
0.666071374	0.681478298	0.25760519	0.285586241	0.283693552	0.299792879	0.4327481	0.445226292
0.636330937	0.646975467	0.05761645	0.069834309	0.29925125	0.320435618	0.440526947	0.455249427
0.285016588	0.281777803	0.09911678	0.118817042	0.309874324	0.334265639	0.447260892	0.4642178
0.305581424	0.307989038	0.09953502	0.118972956	0.322532775	0.350801277	0.467630772	0.490694814
0.316238506	0.321609669	0.11937824	0.142097419	0.345216693	0.379455744	0.480401648	0.506404736
0.331350588	0.340460308	0.16104049	0.188962284	0.358171157	0.395063854	0.510635945	0.544026889
0.353404239	0.367720779	0.16590355	0.193305803	0.390751637	0.43218305	0.52892327	0.562017886
0.362473843	0.378914072	0.21495894	0.243265948	0.413390184	0.451820341	0.516560441	0.547924446
0.396209899	0.416160982	0.24540104	0.271386883	0.380152917	0.420463658		
0.420109276	0.438695526	0.20982833	0.238723023	0.238152693	0.258811181		
0.390775756	0.41025661	0.70565322	0.683287128	0.255374026	0.281651402		
0.178322097	0.182153553	0.71283127	0.68875634	0.264022127	0.292947512		

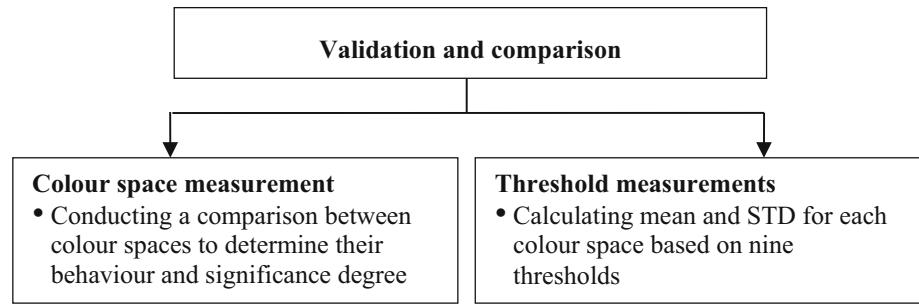
**Fig. 33** Internal and external aggregation ranking

Findings show that the ideal values are 0.8160529 and 0.7867003 for the internal and external aggregations, respectively. The results of the colour space algorithms according to the external and internal aggregations reach averages of  $0.4442933 \pm 0.4827987$  and  $0.432439 \pm$

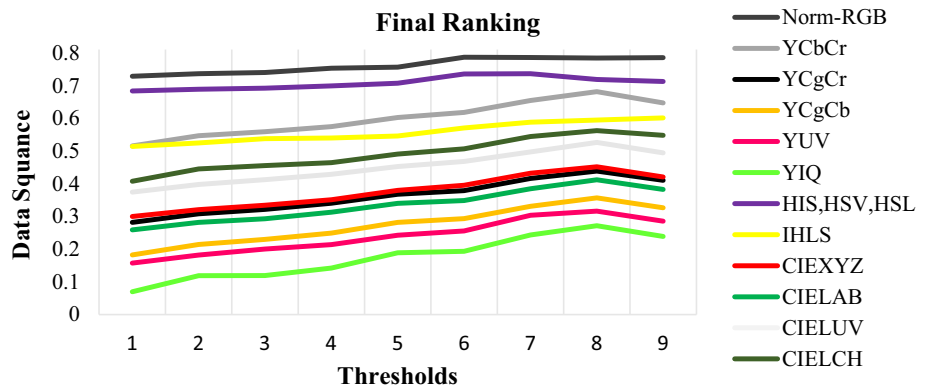
0.4776001, respectively. The outcome of aggregation indicates similarities between the internal and external rankings of various colour space algorithms based on the perspective of the six evaluators. Figure 33 depicts the post-aggregation performance of the tested colour space



**Fig. 34** Overview of the design and implementation of a validation process



**Fig. 35** Colour space measurement



algorithms based on the results of the six external evaluators.

### 4 Validity process

This section provides an insight into the validation and comparison of the different colour spaces obtained for this research. The validation process is an important measure for various empirical researchers in validating the accuracy of the results. Therefore, the results obtained must be validated according to the MCDM techniques. For this procedure, the initial step is the comparison of the different colour spaces based on the results. The statistical measurement method is used to calculate the values for the mean and standard deviation of each threshold value used. Figure 34 shows the overview of the design and implementation of a validation process.

#### 4.1 Colour space measurement

The selection of the appropriate alternative based on the different criteria in skin detection can be implemented using the TOPSIS method. Thus, two categories of results, namely external and internal aggregations, are collected. This study aims to obtain results from external aggregation because it includes all values of comparison between the

criteria and alternatives based on the calculation of averages amongst different ranking values. Figure 35 shows the behaviour of different colour spaces according to the criteria at specific threshold values.

Figure 35 shows that the behaviour of all the colour spaces is relatively identical based on their original values. The behaviour of YIQ starts at the lowest value at threshold 1 and increases gradually until it reaches threshold 2. It stabilises to threshold 3, gradually rises to threshold 5 and stabilises until it reaches threshold 6. It gradually rises to threshold 8 and stops at threshold 9. The behaviour of YUV starts with a gradual rise from threshold 1 to threshold 4. It continues to rise to thresholds 6 and 7 and stabilises until threshold 8 before it declines at threshold 9. The behaviours of YCbCr and CIELAB start from thresholds 1 and 2 and gradually rise to threshold 5. They stabilise at threshold 6, rise to threshold 8 and decline to threshold 9. YCrCb and CIEXYZ show identical behaviours and start a gradual rise from threshold 1 to threshold 5. They stabilise slightly at threshold 6 and gradually rise to threshold 8. Finally, they drop slightly to threshold 9. By contrast, CIELUV shows a distinct behaviour as it exhibits a gradual rise from threshold 1 to threshold 8. It declines slightly to threshold 9. CIELCH rises from threshold 1 to threshold 2 and slightly rises to threshold 5. It stabilises slightly to threshold 6, gradually rises to threshold 8 and drops to threshold 9. IHLS shows a distinct behaviour as it to

**Table 16** Mean and standard deviation for threshold values

	Mean	$\pm$ SD
$\emptyset = 0.5$	0.372021	0.256188
$\emptyset = 0.6$	0.392262	0.249095
$\emptyset = 0.65$	0.400940	0.247265
$\emptyset = 0.7$	0.413236	0.244608
$\emptyset = 0.75$	0.432983	0.234217
$\emptyset = 0.8$	0.446975	0.241581
$\emptyset = 0.85$	0.473835	0.223924
$\emptyset = 0.9$	0.491004	0.210286
$\emptyset = 0.95$	0.468699	0.223906

gradually rise from threshold 1 until threshold 9 at the same level. YCbCr starts from threshold 1 to threshold 2 and slightly rises to threshold 5. It stabilises slightly at threshold 6, gradually rises to threshold 8 and declines to threshold 9. The behaviours of HIS, HSV and HSL start from threshold 1 and rise gradually until they reach threshold 5. Afterwards, they slightly rise to threshold 6, stabilise at threshold 7 and slightly decline slightly until threshold 9. Finally, normalised RGB starts from threshold 1 and gradually rises until threshold 5. It slightly rises to threshold 6 before it stabilises until threshold 9.

Therefore, the behaviour of all the colour spaces varies corresponding to their obtained values from external aggregation. The colour space YIQ yields the lowest value, whereas normalised RGB records the highest value amongst the colour spaces. The rest of the colours come sequentially as shown in Fig. 35.

## 4.2 Threshold measurements

This section presents the calculation of mean and standard deviation value of the threshold values for all colour spaces. Nine threshold values are adopted based on the case study in this research. The threshold values are distributed as 0.5, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9 and 0.95 for each colour space (see Table 16).

Table 16 highlights the results obtained for each threshold value. The results show that the mean value is the lowest with threshold 0.5 and gradually increases with the threshold until it reaches the highest value at the threshold of 0.9 and decline at threshold 0.95. The findings also indicate that the best result is at the threshold of 0.9.

## 5 Conclusion

In this work, the IoT system is adopted to collect data required for analysis, evaluation and comparison, and MCDM techniques are utilised to evaluate and benchmark skin detection approaches. Firstly, different images are collected from the real-time camera in an IoT environment. Processes are applied on the basis of developed skin detection approaches by using multi-agent learning in different colour spaces to create the decision matrix. The decision matrix is evaluated on the basis of different criteria with various skin engines. The performance of multidimensional criteria for skin detection engines is verified by identifying the correlation between different criteria through the Pearson method and by determining their behaviour based on nine thresholds at each colour space. The results confirm that the existing statistical significant differences and differentiated behaviours between criteria ensure that all the available criteria used are needed in the evaluation because multidimensional measurements are necessary. Moreover, the integration between MCDM techniques for the evaluation and benchmarking of skin detection approaches is presented. This research utilises the ML-AHP method based on the pairwise technique, which relies on the differences in evaluator preferences. This approach is the best technique for the identification of the weights of the criteria. The TOPSIS method used is the ideal MCDM technique for the selection of the best alternative based on group decision-making in skin detection approaches. Accordingly, the overall comparison of external and internal aggregation values presented that the normalised RGB at the sixth threshold is the highest value and was the most recommended of all spaces, whereas the YIQ colour space had the lowest value and was the worst case. Moreover, the lowest threshold was obtained at 0.5, whereas the best value was 0.9. Further studies should focus on time complexity algorithms by considering the criteria of the evaluation and benchmarking of real-time IoT skin detectors and defining their procedure objectively.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

## Appendix

See Tables 17, 18, 19, 20, 21, and 22.

**Table 17** First evaluator results to evaluate and benchmark for different colour space algorithms

<i>S</i> −	<i>S</i> +	Rank-1	<i>S</i> −	<i>S</i> +	Rank-1	<i>S</i> −	<i>S</i> +	Rank-1	<i>S</i> −	<i>S</i> +	Rank-1
0.017116	0.023158	0.424983	0.009156	0.022128	0.292674	0.015896	0.020653	0.43492	0.013994	0.017025	0.451157
0.01710	0.021884	0.438591	0.010468	0.021006	0.332587	0.015979	0.018949	0.457475	0.016688	0.014636	0.532751
0.017451	0.021761	0.445044	0.012167	0.019615	0.382829	0.016506	0.018181	0.475855	0.017815	0.01397	0.560475
0.017711	0.01998	0.469907	0.015418	0.017485	0.468584	0.01851	0.015249	0.548308	0.02190	0.01148	0.656041
0.017574	0.019217	0.477674	0.016609	0.016912	0.495476	0.019311	0.016471	0.539681	0.025756	0.010115	0.718008
0.019251	0.01618	0.543337	0.020919	0.014871	0.584491	0.018909	0.018659	0.503319	0.02140	0.011558	0.649348
0.019658	0.016476	0.544033	0.024939	0.014025	0.640054	0.019207	0.019209	0.499965	0.010957	0.021933	0.333146
0.020242	0.016829	0.546028	0.020382	0.015042	0.575386	0.013934	0.021122	0.397477	0.012632	0.019034	0.398909
0.020012	0.016694	0.54520	0.006275	0.024552	0.20355	0.014446	0.019271	0.428442	0.013728	0.017568	0.438654
0.012038	0.022626	0.347277	0.008922	0.022309	0.285672	0.015226	0.01800	0.458223	0.015115	0.016235	0.482135
0.013778	0.018305	0.42946	0.01020	0.021172	0.325073	0.015469	0.017288	0.472242	0.017511	0.013695	0.56115
0.014687	0.016925	0.46460	0.01190	0.020192	0.370863	0.015868	0.016178	0.495166	0.01860	0.012329	0.601348
0.015915	0.015169	0.51200	0.014591	0.018471	0.441309	0.017956	0.013951	0.562754	0.022321	0.010186	0.686651
0.01823	0.012524	0.592768	0.016206	0.017681	0.478228	0.019364	0.01190	0.619292	0.026254	0.008256	0.760767
0.019224	0.011506	0.625583	0.020438	0.015237	0.572888	0.020518	0.011857	0.633375	0.021894	0.010309	0.67988
0.023195	0.008182	0.739238	0.02330	0.01499	0.608563	0.020138	0.011537	0.635763	0.012113	0.022977	0.345187
0.026927	0.00640	0.808013	0.019744	0.016251	0.548517	0.010175	0.022137	0.31490	0.014357	0.01766	0.448432
0.022262	0.00880	0.716689	0.004807	0.026663	0.152757	0.011829	0.01950	0.377605	0.015118	0.016579	0.476947
0.007669	0.023758	0.244021	0.008494	0.023973	0.261619	0.012976	0.018135	0.417086	0.015877	0.015522	0.505644
0.009927	0.021132	0.319624	0.008573	0.024146	0.262031	0.014384	0.016539	0.465159	0.018085	0.012969	0.582367
0.011158	0.019982	0.358322	0.01036	0.022762	0.312786	0.016982	0.014108	0.546218	0.01915	0.011923	0.61630
0.01287	0.018686	0.407856	0.014269	0.020216	0.413765	0.01800	0.013178	0.577385	0.022987	0.008582	0.728141
0.015811	0.01652	0.489042	0.014818	0.020313	0.42180	0.022239	0.010351	0.682394	0.026526	0.007128	0.78820
0.01690	0.015562	0.520636	0.019962	0.018171	0.52348	0.025965	0.009292	0.736451	0.025135	0.008226	0.753416
0.021132	0.013556	0.60920	0.023494	0.017284	0.576149	0.02120	0.011007	0.658184			
0.025116	0.012542	0.666941	0.019389	0.01821	0.515676	0.00960	0.022484	0.299276			
0.020591	0.013824	0.598319	0.015769	0.022035	0.417118	0.011454	0.019838	0.366046			
0.00645	0.0246	0.207729	0.015712	0.020789	0.430455	0.01246	0.018731	0.399483			

**Table 18** Second evaluator results to evaluate and benchmark for different colour space algorithms

<i>S</i> −	<i>S</i> +	Rank-2	<i>S</i> −	<i>S</i> +	Rank-2	<i>S</i> −	<i>S</i> +	Rank-2	<i>S</i> −	<i>S</i> +	Rank-2
0.046322	0.003036	0.938494	0.006707	0.041529	0.139038	0.03947	0.007516	0.84004	0.005623	0.041159	0.120201
0.04632	0.002893	0.941223	0.006735	0.041519	0.139575	0.039461	0.007442	0.841341	0.005724	0.041144	0.122125
0.046327	0.002895	0.941184	0.006777	0.041507	0.140361	0.039465	0.007413	0.841868	0.005775	0.04114	0.123097
0.046325	0.002657	0.945756	0.006877	0.041492	0.142186	0.039472	0.007288	0.844138	0.005991	0.041128	0.127153
0.046316	0.002582	0.947204	0.006926	0.041486	0.143066	0.039485	0.00732	0.843613	0.006256	0.041122	0.132047
0.046321	0.002157	0.955508	0.00711	0.041473	0.146338	0.039479	0.007412	0.841935	0.005968	0.041128	0.126724
0.046322	0.002178	0.95509	0.007328	0.04147	0.150167	0.039471	0.007428	0.841613	0.015611	0.030939	0.335358
0.046319	0.002185	0.954958	0.007084	0.041475	0.145892	0.024163	0.022414	0.518773	0.015627	0.030907	0.335819
0.046321	0.002185	0.954949	0.00592	0.041274	0.125443	0.024163	0.022388	0.519067	0.015641	0.030892	0.336133
0.029496	0.017708	0.624859	0.005954	0.041257	0.126122	0.024174	0.022368	0.519393	0.015662	0.030877	0.336534
0.029505	0.017627	0.626008	0.005989	0.041243	0.126806	0.024169	0.022359	0.519447	0.015693	0.030859	0.337112
0.029511	0.017603	0.62637	0.00603	0.041238	0.12757	0.024168	0.022349	0.519546	0.015715	0.03085	0.337478
0.029518	0.017579	0.626756	0.006121	0.041225	0.12929	0.024185	0.022323	0.520024	0.015791	0.030837	0.338656
0.029537	0.017543	0.627382	0.006176	0.041223	0.13029	0.024199	0.022303	0.520388	0.015894	0.030829	0.34018
0.029552	0.017526	0.62772	0.006396	0.041195	0.134395	0.024211	0.022298	0.520562	0.01578	0.030838	0.338502
0.029594	0.017499	0.628418	0.006548	0.041204	0.137135	0.02422	0.022295	0.520698	0.015725	0.031081	0.335961
0.02965	0.017485	0.629038	0.006347	0.041215	0.133439	0.00878	0.037766	0.188638	0.015748	0.031025	0.336686
0.029583	0.017503	0.628276	0.00041	0.046401	0.008748	0.008802	0.037742	0.189106	0.015757	0.031014	0.336899
0.016222	0.031263	0.341627	0.000891	0.046374	0.018858	0.008824	0.03773	0.189543	0.015765	0.031005	0.337073
0.016238	0.031231	0.342076	0.000897	0.046375	0.018981	0.008853	0.037718	0.190102	0.015799	0.030985	0.337707
0.01625	0.031218	0.342327	0.001144	0.046364	0.02407	0.008923	0.037701	0.191374	0.015828	0.030976	0.338184
0.016269	0.031202	0.342717	0.001658	0.046345	0.034534	0.008968	0.037692	0.192201	0.015907	0.030958	0.339418
0.016308	0.031185	0.34337	0.001745	0.046344	0.036295	0.009116	0.037678	0.194811	0.01599	0.030955	0.340618
0.016324	0.03118	0.343634	0.002402	0.046331	0.049292	0.009274	0.037676	0.197537	0.015948	0.030958	0.339995
0.016407	0.031159	0.344934	0.002908	0.046327	0.059059	0.009069	0.037683	0.193974			
0.016499	0.031158	0.34621	0.002346	0.046333	0.048198	0.005497	0.041204	0.117704			
0.016395	0.031163	0.344729	0.039471	0.007568	0.839113	0.005538	0.041181	0.118539			
0.006664	0.041555	0.138195	0.039466	0.007518	0.839996	0.005566	0.041171	0.119093			

**Table 19** Third evaluator results to evaluate and benchmark for different colour space algorithms

S−	S+	Rank-3	S−	S+	Rank-3	S−	S+	Rank-3	S−	S+	Rank-3
0.028304	0.006507	0.813078	0.005219	0.025144	0.171881	0.0246	0.007016	0.778097	0.008446	0.023141	0.267388
0.028245	0.006159	0.820989	0.005423	0.025055	0.17793	0.024306	0.00678	0.781895	0.008637	0.023087	0.272251
0.028451	0.006115	0.823085	0.005708	0.024953	0.186161	0.024377	0.006586	0.787306	0.008699	0.023102	0.273552
0.028361	0.005617	0.834675	0.006236	0.02488	0.200413	0.024386	0.006033	0.801676	0.009383	0.023003	0.289712
0.028111	0.005446	0.83772	0.006496	0.024822	0.207419	0.024801	0.006057	0.803702	0.010199	0.022915	0.308003
0.028111	0.004621	0.858828	0.00741	0.024719	0.230638	0.02473	0.006556	0.790455	0.009334	0.022987	0.288791
0.028118	0.004702	0.856745	0.00831	0.024707	0.25168	0.024495	0.006804	0.782609	0.011865	0.018274	0.393676
0.028025	0.004843	0.852661	0.00728	0.02474	0.227355	0.017179	0.013419	0.561448	0.011953	0.018035	0.398584
0.028095	0.004772	0.8548	0.004422	0.025342	0.148573	0.0171	0.013242	0.563579	0.012066	0.017908	0.402551
0.018127	0.012275	0.596238	0.004497	0.025409	0.150379	0.017264	0.013056	0.569393	0.012232	0.017787	0.407481
0.018193	0.011721	0.608174	0.004846	0.025194	0.16131	0.017076	0.013062	0.566609	0.012254	0.017743	0.408508
0.01822	0.011583	0.611344	0.004954	0.025324	0.16362	0.016959	0.013015	0.56579	0.012435	0.017633	0.413567
0.018229	0.011459	0.614026	0.005464	0.025248	0.177912	0.017102	0.012817	0.571606	0.012798	0.017588	0.421174
0.018342	0.011261	0.619605	0.005715	0.025347	0.183993	0.017081	0.012725	0.573074	0.013423	0.017477	0.434404
0.01856	0.011031	0.627212	0.006937	0.024839	0.218305	0.017237	0.01269	0.575956	0.012723	0.017605	0.419508
0.018763	0.010957	0.631335	0.00742	0.025152	0.227804	0.017523	0.012517	0.583325	0.013226	0.017923	0.424595
0.01923	0.010774	0.640908	0.006477	0.025373	0.203358	0.009257	0.021659	0.299425	0.013334	0.017465	0.432927
0.018701	0.010965	0.630396	0.00137	0.028812	0.045399	0.00929	0.021504	0.301687	0.013377	0.017389	0.43481
0.00952	0.020394	0.318237	0.002365	0.028652	0.076255	0.009385	0.021419	0.304672	0.013361	0.017342	0.435172
0.009656	0.020195	0.323474	0.002378	0.028691	0.076525	0.009498	0.021336	0.308049	0.013517	0.0172	0.440052
0.009748	0.020118	0.326404	0.002875	0.028616	0.091302	0.009784	0.021216	0.315615	0.013837	0.017062	0.447815
0.009933	0.019988	0.331974	0.003973	0.028465	0.122466	0.010164	0.021065	0.325455	0.014168	0.01695	0.455303
0.010164	0.019976	0.337216	0.004119	0.028499	0.126284	0.010699	0.020988	0.337661	0.014417	0.016996	0.458954
0.010281	0.019948	0.340114	0.005566	0.028442	0.163675	0.011207	0.021017	0.347787	0.014122	0.017077	0.452649
0.010943	0.019722	0.356856	0.006562	0.028443	0.187464	0.010388	0.021081	0.330107			
0.011514	0.019796	0.367742	0.005406	0.028489	0.159485	0.008008	0.023488	0.254254			
0.010813	0.019815	0.353047	0.024624	0.007324	0.770762	0.008118	0.023326	0.258177			
0.004862	0.025351	0.160935	0.024486	0.007092	0.775421	0.008183	0.023272	0.260149			

**Table 20** Fourth evaluator results to evaluate and benchmark for different colour space algorithms

<i>S</i> −	<i>S</i> +	Rank-4	<i>S</i> −	<i>S</i> +	Rank-4	<i>S</i> −	<i>S</i> +	Rank-4	<i>S</i> −	<i>S</i> +	Rank-4
0.015111	0.018951	0.443638	0.007501	0.018849	0.284675	0.013887	0.016931	0.450615	0.011582	0.01459	0.442541
0.015088	0.017913	0.45719	0.00857	0.017968	0.322941	0.013908	0.015543	0.47224	0.013751	0.012744	0.519004
0.015383	0.017816	0.463347	0.009955	0.016882	0.370936	0.014315	0.01492	0.489662	0.014664	0.012234	0.54518
0.015563	0.016358	0.48754	0.012607	0.015239	0.452735	0.015845	0.012523	0.558556	0.017979	0.010362	0.634391
0.015422	0.01574	0.494891	0.013585	0.014791	0.478745	0.016501	0.01351	0.549828	0.021133	0.009358	0.69308
0.016683	0.013252	0.557317	0.017104	0.013246	0.563567	0.016176	0.015301	0.513903	0.017581	0.010417	0.627937
0.016989	0.013492	0.557365	0.020397	0.01262	0.617783	0.01637	0.015748	0.509683	0.009343	0.018221	0.338946
0.017415	0.013775	0.558346	0.016666	0.013376	0.554754	0.012002	0.01741	0.408063	0.010654	0.015891	0.401356
0.017252	0.013667	0.557967	0.005215	0.02063	0.201777	0.012386	0.015913	0.437684	0.011525	0.014716	0.439191
0.010599	0.018599	0.362996	0.007331	0.018866	0.279849	0.013004	0.01488	0.466357	0.012633	0.013645	0.480741
0.011922	0.01509	0.441361	0.008374	0.017952	0.318071	0.013174	0.014304	0.479434	0.014541	0.011646	0.555276
0.012622	0.01397	0.474639	0.009754	0.017197	0.361924	0.013471	0.013417	0.501006	0.015418	0.01058	0.593044
0.01357	0.012554	0.519457	0.011946	0.01585	0.429776	0.01511	0.011624	0.565188	0.018425	0.008924	0.673713
0.015388	0.01042	0.596245	0.013256	0.015256	0.464927	0.016221	0.009986	0.618953	0.02162	0.007478	0.743004
0.016189	0.009587	0.628063	0.016729	0.013317	0.556794	0.017137	0.009939	0.632929	0.018078	0.009021	0.667119
0.019347	0.006955	0.735581	0.019068	0.013177	0.591348	0.016865	0.009675	0.635454	0.010315	0.019087	0.350825
0.022355	0.005556	0.800942	0.016153	0.014161	0.532846	0.008592	0.018534	0.31673	0.012075	0.01483	0.448789
0.018602	0.007439	0.714324	0.003902	0.022546	0.147521	0.009892	0.016433	0.375768	0.012677	0.013969	0.475746
0.006627	0.019775	0.250993	0.006915	0.02042	0.252985	0.010807	0.015351	0.413134	0.013274	0.013133	0.502677
0.008386	0.017668	0.321872	0.00698	0.020556	0.253498	0.011933	0.014096	0.458438	0.015039	0.011126	0.57478
0.009361	0.016749	0.358514	0.008443	0.019472	0.302456	0.014027	0.0122	0.534834	0.015911	0.010301	0.607016
0.01073	0.015708	0.405842	0.01164	0.017493	0.399538	0.014868	0.011466	0.564588	0.019	0.007769	0.709775
0.01309	0.014004	0.483137	0.012095	0.017563	0.407812	0.018298	0.009331	0.662265	0.021859	0.006734	0.764473
0.013968	0.013259	0.513015	0.016296	0.015927	0.505731	0.021328	0.008568	0.713395	0.020726	0.007523	0.733699
0.017405	0.011672	0.598589	0.019195	0.015259	0.557114	0.017445	0.009832	0.639546			
0.020642	0.010912	0.654178	0.015832	0.015965	0.497913	0.008075	0.018903	0.299324			
0.016963	0.01189	0.587903	0.013795	0.018052	0.433171	0.009542	0.016803	0.362195			
0.005306	0.020805	0.203192	0.013733	0.01704	0.446267	0.010345	0.015928	0.393754			

**Table 21** Fifth evaluator results to evaluate and benchmark for different colour space algorithms

S−	S+	Rank-5	S−	S+	Rank-5	S−	S+	Rank-5	S−	S+	Rank-5	S−	S+	Rank-5
0.054019	0.0025915	0.9542217	0.0098101	0.0447446	0.1798211	0.0470765	0.007376	0.8645417	0.0073579	0.0472347	0.1347784			
0.0540137	0.0024486	0.9566332	0.0098271	0.0447366	0.1801032	0.04705	0.0073424	0.8650113	0.0073887	0.047231	0.1352752			
0.0540322	0.0024267	0.9570176	0.0098528	0.0447272	0.1805207	0.0470563	0.0073104	0.8655346	0.0073927	0.0472329	0.135334			
0.0540238	0.002226	0.9604259	0.0099019	0.0447212	0.1812766	0.0470563	0.0072306	0.8668082	0.0075297	0.0472249	0.1375175			
0.0540015	0.0021466	0.9617695	0.0099243	0.0447163	0.1816288	0.0470934	0.0072243	0.8669996	0.0076909	0.0472182	0.1400657			
0.0540009	0.0017957	0.9678173	0.0100296	0.0447069	0.1832336	0.047087	0.0072866	0.8659899	0.0075189	0.0472237	0.1373498			
0.0540015	0.0018107	0.967558	0.0101359	0.0447064	0.1848185	0.0470658	0.0073187	0.8654273	0.0173629	0.036873	0.3201372			
0.053993	0.0018468	0.9669265	0.0100137	0.0447089	0.1829901	0.0274547	0.0267998	0.5060358	0.0173712	0.0368544	0.3203501			
0.0539994	0.0018264	0.9672842	0.0043737	0.0501777	0.0801758	0.0274457	0.0267858	0.5060842	0.0173833	0.0368445	0.3205608			
0.0311852	0.0231853	0.5735684	0.0043788	0.0501845	0.0802523	0.0274627	0.026771	0.5063769	0.0174008	0.0368353	0.3208343			
0.0311908	0.0231393	0.5740978	0.0044378	0.0501665	0.0812721	0.0274421	0.0267722	0.5061787	0.0173998	0.0368326	0.3208375			
0.0311927	0.0231285	0.5742275	0.00445	0.0501783	0.0814597	0.0274297	0.0267684	0.5061012	0.0174225	0.0368236	0.321175			
0.0311931	0.0231186	0.5743342	0.0045385	0.0501725	0.0829539	0.0274443	0.0267518	0.5063891	0.0174581	0.036821	0.3216355			
0.0312028	0.0231036	0.574569	0.0045929	0.0501808	0.0838522	0.0274407	0.0267459	0.5064108	0.0175339	0.0368124	0.3226323			
0.0312239	0.0230855	0.5749254	0.0048378	0.0501386	0.0879979	0.0274576	0.0267419	0.5066025	0.0174501	0.0368224	0.3215271			
0.0312417	0.0230809	0.5751138	0.0049455	0.0501654	0.0897373	0.0274883	0.0267289	0.5070031	0.0202366	0.0341263	0.3722502			
0.0312864	0.0230672	0.5756092	0.0047357	0.0501841	0.0862291	0.0105808	0.0437882	0.1946113	0.0202466	0.034088	0.3726287			
0.0312359	0.0230814	0.5750634	0.0005089	0.0540611	0.0093255	0.0105828	0.0437763	0.1946838	0.0202507	0.0340819	0.3727178			
0.0146269	0.0398701	0.2683979	0.0008934	0.054048	0.0162615	0.0105949	0.0437697	0.1948857	0.0202482	0.0340783	0.3727127			
0.0146397	0.0398541	0.2686494	0.0008902	0.0540516	0.016203	0.0106105	0.0437632	0.1951401	0.0202635	0.034067	0.3729672			
0.0146487	0.039848	0.2688	0.0010924	0.0540453	0.0198125	0.0106501	0.0437541	0.1957589	0.0202988	0.0340557	0.3734524			
0.0146671	0.0398376	0.2690977	0.0015483	0.0540323	0.0278563	0.0107078	0.0437421	0.1966542	0.0203334	0.0340472	0.3739092			
0.0146898	0.0398376	0.2694019	0.0015725	0.0540359	0.028278	0.0107886	0.0437365	0.1978652	0.0203575	0.0340518	0.3741548			
0.0147047	0.0398351	0.2696143	0.0022029	0.0540308	0.0391743	0.0108655	0.0437393	0.198985	0.0203204	0.0340585	0.3736814			
0.014775	0.0398169	0.2706449	0.0025906	0.0540316	0.0457526	0.0107376	0.0437441	0.1970862						
0.014844	0.0398235	0.271532	0.0021269	0.0540352	0.03787	0.0072835	0.047262	0.1335302						
0.0147603	0.0398247	0.2704089	0.0470786	0.0074261	0.8637537	0.0072996	0.0472494	0.1338167						
0.0097826	0.0447633	0.1793461	0.0470662	0.0073895	0.8643028	0.0073088	0.0472454	0.1339735						

**Table 22** Sixth evaluator results to evaluate and benchmark for different colour space algorithms

S−	S+	Rank-6	S−	S+	Rank-6	S−	S+	Rank-6	S−	S+	Rank-6
0.031632	0.008088	0.796363	0.007272	0.026229	0.217082	0.028762	0.007952	0.7834	0.015655	0.018275	0.461397
0.031398	0.007668	0.803709	0.007536	0.02606	0.224322	0.027605	0.008028	0.774706	0.015516	0.018449	0.456813
0.032177	0.007584	0.809254	0.00789	0.025871	0.233699	0.027816	0.007707	0.783048	0.015386	0.018656	0.451973
0.031787	0.006977	0.820016	0.008367	0.025914	0.244064	0.027565	0.007165	0.793706	0.015958	0.018503	0.463066
0.030792	0.00686	0.817802	0.008729	0.025753	0.253146	0.029071	0.006634	0.814205	0.016851	0.018187	0.480931
0.030567	0.005936	0.837394	0.009788	0.0256	0.276591	0.028865	0.007386	0.796258	0.015985	0.0184	0.464887
0.030541	0.006053	0.834587	0.010801	0.025658	0.296246	0.027909	0.00807	0.775695	0.017921	0.016275	0.524064
0.030071	0.006363	0.82536	0.009609	0.02567	0.272383	0.024456	0.010913	0.691458	0.017936	0.015928	0.529646
0.030397	0.006185	0.830923	0.006155	0.027328	0.183836	0.024137	0.010733	0.692188	0.018092	0.01567	0.535864
0.020311	0.014148	0.589421	0.005645	0.028051	0.167534	0.024549	0.010211	0.706237	0.018342	0.01539	0.543763
0.020328	0.013471	0.60144	0.006401	0.027377	0.189493	0.023918	0.010532	0.694284	0.017911	0.01573	0.532408
0.02029	0.013355	0.603059	0.006096	0.028083	0.178346	0.023479	0.01068	0.687337	0.018214	0.01543	0.541366
0.020115	0.013365	0.600802	0.006687	0.028053	0.192497	0.023602	0.01034	0.695354	0.018384	0.015569	0.541442
0.020125	0.01325	0.603009	0.006768	0.028567	0.191548	0.023273	0.010425	0.690643	0.019164	0.015263	0.556647
0.020801	0.012579	0.623149	0.008928	0.026836	0.249637	0.023565	0.010267	0.696533	0.018266	0.015635	0.538809
0.020681	0.012888	0.616076	0.009043	0.028147	0.243145	0.024568	0.009336	0.724621	0.021653	0.013477	0.616369
0.021545	0.012418	0.634359	0.007634	0.028868	0.209128	0.016595	0.01766	0.484456	0.021641	0.012607	0.631895
0.020667	0.012823	0.617105	0.00188	0.032141	0.055255	0.016431	0.017533	0.483765	0.021645	0.012475	0.634377
0.008945	0.024508	0.26739	0.003046	0.031997	0.086925	0.016476	0.017407	0.486272	0.021475	0.012503	0.632028
0.0091	0.024326	0.272238	0.003046	0.032126	0.086599	0.016501	0.017318	0.487921	0.021512	0.012296	0.636295
0.009217	0.024265	0.275291	0.003648	0.032061	0.102158	0.016683	0.017161	0.492934	0.022203	0.01166	0.655663
0.009589	0.024025	0.285276	0.004998	0.031853	0.135615	0.01744	0.016483	0.5141	0.022299	0.01161	0.657615
0.009666	0.024349	0.284157	0.005178	0.031976	0.139367	0.017747	0.016507	0.518102	0.022106	0.01213	0.645706
0.009792	0.024388	0.28647	0.006952	0.032051	0.178243	0.017979	0.016812	0.516767	0.02163	0.012481	0.634107
0.011049	0.023836	0.316738	0.008182	0.032166	0.202783	0.017203	0.016938	0.503884			
0.011699	0.024235	0.32557	0.006751	0.032228	0.173196	0.015369	0.018877	0.448779			
0.010703	0.024145	0.307133	0.02887	0.008343	0.775806	0.015344	0.018668	0.451135			
0.006799	0.026607	0.203525	0.028342	0.008177	0.776096	0.015321	0.018632	0.451233			

## References

- Zaidan A, Zaidan B et al (2018) A review on intelligent process for smart home applications based on IoT: coherent taxonomy, motivation, open challenges, and recommendations. *Artif Intell Rev.* <https://doi.org/10.1007/s10462-018-9648-9>
- Tan J, Koo SG (2014) A survey of technologies in internet of things. In: 2014 IEEE international conference on distributed computing in sensor systems. IEEE
- Gia TN et al (2018) Energy efficient wearable sensor node for IoT-based fall detection systems. *Microprocess Microsyst* 56:34–46
- Ishii H et al (2016) An early detection system for dementia using the M2M/IoT platform. *Procedia Comput Sci* 96:1332–1340
- Kapoor A et al (2016) Implementation of IoT (Internet of Things) and image processing in smart agriculture. In: 2016 international conference on computation system and information technology for sustainable solutions (CSITSS). IEEE
- Hu P et al (2018) A unified face identification and resolution scheme using cloud computing in Internet of Things. *Future Gener Comput Syst* 81:582–592
- Madeira R, Nunes L (2016) A machine learning approach for indirect human presence detection using IoT devices. In: 2016 eleventh international conference on digital information management (ICDIM). IEEE
- Talal M et al (2019) Smart home-based IoT for real-time and secure remote health monitoring of triage and priority system using body sensors: multi-driven systematic review. *J Med Syst* 43(3):42
- Albahri O et al (2018) Systematic review of real-time remote health monitoring system in triage and priority-based sensor technology: taxonomy, open challenges, motivation and recommendations. *J Med Syst* 42(5):80
- Zaidan AA et al (2018) A survey on communication components for IoT-based technologies in smart homes. *Telecommun Syst* 69(1):1–25
- Alaa M et al (2017) A review of smart home applications based on Internet of Things. *J Netw Comput Appl* 97:48–65
- Rupani A et al (2017) A robust technique for image processing based on interfacing of Raspberry-Pi and FPGA using IoT. In: 2017 international conference on computer, communications and electronics (CompTelx). IEEE
- Liao M-S et al (2017) On precisely relating the growth of Phalaenopsis leaves to greenhouse environmental factors by using an IoT-based monitoring system. *Comput Electron Agric* 136:125–139



14. Marimuthu R et al (2017) Driver fatigue detection using image processing and accident prevention. *Int J Pure Appl Math* 116(11):91–99
15. Lee H (2017) Framework and development of fault detection classification using IoT device and cloud environment. *J Manuf Syst* 43:257–270
16. Dinesh M, Sudhaman K (2016) Real time intelligent image processing system with high speed secured Internet of Things: image processor with IOT. In: 2016 international conference on information communication and embedded systems (ICICES). IEEE
17. Gu J, Huayu Y (2015) Real-time image collection and processing system design. In: 2015 fifth international conference on instrumentation and measurement, computer, communication and control (IMCCC). IEEE
18. Kongurusa N, Chumuang N, Ketcham M (2017) Real-time intrusion—detecting and alert system by image processing techniques. In: 2017 10th international conference on Ubi-media computing and workshops (Ubi-Media). IEEE
19. Wu Q et al (2014) Cognitive internet of things: a new paradigm beyond connection. *IEEE Internet Things J* 1(2):129–143
20. Nesa N, Ghosh T, Banerjee I (2018) Non-parametric sequence-based learning approach for outlier detection in IoT. *Future Gener Comput Syst* 82:412–421
21. Beikhhakhan Y et al (2015) The application of ISM model in evaluating agile suppliers selection criteria and ranking suppliers using fuzzy TOPSIS-AHP methods. *Expert Syst Appl* 42(15–16):6224–6236
22. Hasan S, Curry E (2015) Thingsonomy: tackling variety in internet of things events. *IEEE Internet Comput* 19(2):10–18
23. Jones MJ, Rehg JM (1999) Statistical color models with application to skin detection. In: Proceedings. 1999 IEEE computer society conference on computer vision and pattern recognition (Cat. No. PR00149)
24. Phung SL, Bouzerdoum A, Chai D (2005) Skin segmentation using color pixel classification: analysis and comparison. *IEEE Trans Pattern Anal Mach Intell* 1:148–154
25. Gamage N, Akmeliawati R, Chow KY (2009) Towards robust skin colour detection and tracking. In: 2009 IEEE instrumentation and measurement technology conference. IEEE
26. Taqa AY, Jalab HA (2010) Increasing the reliability of skin detectors. *Sci Res Essays* 5(17):2480–2490
27. Huang L et al (2015) Robust skin detection in real-world images. *J Vis Commun Image Represent* 29:147–152
28. Jensch D, Mohr D, Zachmann G (2015) A comparative evaluation of three skin color detection approaches. *J Virtual Real Broadcast* 12(1):6
29. Sanmiguel JC, Suja S (2013) Skin detection by dual maximization of detectors agreement for video monitoring. *Pattern Recognit Lett* 34(16):2102–2109
30. Fernandes BJT, Cavalcanti GD, Ren TI (2013) Lateral inhibition pyramidal neural network for image classification. *IEEE Trans Cybern* 43(6):2082–2092
31. Kawulok M (2013) Fast propagation-based skin regions segmentation in color images. In: 2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG). IEEE
32. Yas QM et al (2018) Comprehensive insights into evaluation and benchmarking of real-time skin detectors: review, open issues & challenges, and recommended solutions. *Measurement* 114:243–260
33. Zaidan A et al (2018) A review on smartphone skin cancer diagnosis apps in evaluation and benchmarking: coherent taxonomy, open issues and recommendation pathway solution. *Health Technol* 8(4):223–238
34. Yas QM et al (2018) A systematic review on smartphone skin cancer apps: coherent taxonomy, motivations, open challenges and recommendations, and new research direction. *J Circuits Syst Comput* 27(05):1830003
35. AlSattar H et al (2018) MOGSABAT: a metaheuristic hybrid algorithm for solving multi-objective optimisation problems. *Neural Comput Appl*. <https://doi.org/10.1007/s00521-018-3808-3>
36. Enaizan O et al (2018) Electronic medical record systems: decision support examination framework for individual, security and privacy concerns using multi-perspective analysis. *Health Technol*. <https://doi.org/10.1007/s12553-018-0278-7>
37. Hwang C-L, Yoon K (2012) Multiple attribute decision making: methods and applications a state-of-the-art survey, vol 186. Springer, Berlin
38. Khatari M et al (2019) Multi-criteria evaluation and benchmarking for active queue management methods: open issues, challenges and recommended pathway solutions. *Int J Inf Technol Decis Mak*. <https://doi.org/10.1142/S0219622019300039>
39. Almahdi EM et al (2019) Mobile-based patient monitoring systems: a prioritisation framework using multi-criteria decision-making techniques. *J Med Syst* 43(7):219. <https://doi.org/10.1007/s10916-019-1339-9>
40. Almahdi E et al (2019) Mobile patient monitoring systems from a benchmarking aspect: challenges, open issues and recommended solutions. *J Med Syst* 43(7):207
41. Alsalem M et al (2019) Multiclass benchmarking framework for automated acute leukaemia detection and classification based on BWM and Group-VIKOR. *J Med Syst* 43(7):212
42. Albahri A et al (2019) Fault-tolerant mHealth framework in the context of IoT based real-time wearable health data sensor. *IEEE Access* 7(50052):50080
43. Zaidan A et al (2015) Evaluation and selection of open-source EMR software packages based on integrated AHP and TOPSIS. *J Biomed Inform* 53:390–404
44. Zaidan A et al (2015) Multi-criteria analysis for OS-EMR software selection problem: a comparative study. *Decis Support Syst* 78:15–27
45. Mohammed KI, Zaidan AA, Zaidan BB, Albahri OS, Alsalem MA, Albahri AS, Hadi A, Hashim M (2019) Real-time remote-health monitoring systems: a review on patients prioritisation for multiple-chronic diseases, taxonomy analysis, concerns and solution procedure. *J Med Syst* 43(7):223
46. Zaidan AA, Zaidan BB, Taqa YA, Sami MK, Alam GM, Jalab AH (2010) Novel multi-cover steganography using remote sensing image and general recursion neural cryptosystem. *Int J Phys Sci* 5(11):1776–1786
47. Hodges S et al (2012) Prototyping connected devices for the internet of things. *Computer* 46(2):26–34
48. Satyanarayanan M et al (2015) Edge analytics in the internet of things. *IEEE Pervasive Comput* 14(2):24–31
49. Wehner P, Piberger C, Göhringer D (2014) Using JSON to manage communication between services in the Internet of Things. In: 2014 9th international symposium on reconfigurable and communication-centric systems-on-chip (ReCoSoC). IEEE
50. Atzori L, Iera A, Morabito G (2010) The Internet of Things: a survey. *Comput Netw* 54(15):2787–2805
51. Domingo MC (2012) An overview of the Internet of Things for people with disabilities. *J Netw Comput Appl* 35(2):584–596
52. Posse TAR (2014) A software defined networking architecture for secure routing
53. Want R, Schilit BN, Jenson S (2015) Enabling the internet of things. *Computer* 1:28–35

54. Zaidan A et al (2014) Image skin segmentation based on multi-agent learning Bayesian and neural network. *Eng Appl Artif Intell* 32:136–150
55. Mircea I-G et al (2012) An evaluation of color spaces used in skin color detection. *Stud Univ Babeş-Bolyai Inform* 57(3):24–34
56. Yas QM et al (2017) Towards on develop a framework for the evaluation and benchmarking of skin detectors based on artificial intelligent models using multi-criteria decision-making techniques. *Int J Pattern Recognit Artif Intell* 31(03):1759002
57. Zaidan AA (2013) Anti-pornography algorithm based on multi-agent learning in skin detector and pornography classifier. Multimedia University (Malaysia)
58. Zaidan A et al (2010) A novel hybrid module of skin detector using grouping histogram technique for Bayesian method and segment adjacent-nested technique for neural network. *Int J Phys Sci* 5(16):2471–2492
59. Zaidan A et al (2010) Increase reliability for skin detector using backpropagation neural network and heuristic rules based on YCbCr. *Sci Res Essays* 5(19):2931–2946
60. Zaidan A et al (2014) On the multi-agent learning neural and Bayesian methods in skin detector and pornography classifier: an automated anti-pornography system. *Neurocomputing* 131:397–418
61. Zaidan A et al (2010) A new hybrid module for skin detector using fuzzy inference system structure and explicit rules. *Int J Phys Sci* 5(13):2084–2097
62. Zaidan A et al (2013) An automated anti-pornography system using a skin detector based on artificial intelligence: a review. *Int J Pattern Recognit Artif Intell* 27(04):1350012
63. Zaidan A et al (2014) A four-phases methodology to propose anti-pornography system based on neural and Bayesian methods of artificial intelligence. *Int J Pattern Recognit Artif Intell* 28(01):1459001
64. Powers DM (2013) A computationally and cognitively plausible model of supervised and unsupervised learning. In: *International conference on brain inspired cognitive systems*. Springer
65. Rupanagudi SR et al (2015) A novel cloud computing based smart farming system for early detection of borer insects in tomatoes. In: *2015 international conference on communication, information & computing technology (ICCICT)*. IEEE
66. Al-Mohair HK, Mohamad-Saleh J, Suandi SA (2014) Color space selection for human skin detection using color-texture features and neural networks. In: *2014 international conference on computer and information sciences (ICCOINS)*. IEEE
67. Yang J, Lu W, Waibel A (1998) Skin-color modeling and adaptation. In: *Asian conference on computer vision*. Springer
68. Wang YH, Wu IC et al (2009) Achieving high and consistent rendering performance of Java AWT/Swing on multiple platforms. *Softw Pract Exp* 39(7):701–736
69. Zhang S et al (1998) Calmodulin mediates calcium-dependent inactivation of *N*-methyl-D-aspartate receptors. *Neuron* 21(2):443–453
70. Chai D, Bouzerdoum A (2000) A Bayesian approach to skin color classification in YCbCr color space. In: *2000 TENCON Proceedings. Intelligent systems and technologies for the New Millennium (Cat. No. 00CH37119)*. IEEE
71. Daithankar MV, Karande KJ, Harale AD (2014) Analysis of skin color models for face detection. In: *2014 international conference on communication and signal processing*. IEEE
72. Chaves-González JM et al (2010) Detecting skin in face recognition systems: a colour spaces study. *Digit Signal Proc* 20(3):806–823
73. Zhengzhen Z, Yuexiang S (2009) Skin color detecting unite YCgCb color space with YCgCr color space. In: *2009 international conference on image analysis and signal processing*. IEEE
74. Abadpour A, Kasaei S (2005) Pixel-based skin detection for pornography filtering. *Iran J Electr Electron Eng* 1(3):21–41
75. Yang J et al (2004) Adaptive skin detection using multiple cues. In: *2004 international conference on image processing, 2004. ICIP'04*. IEEE
76. Ma Z, Leijon A (2010) Human skin color detection in RGB space with Bayesian estimation of beta mixture models. In: *2010 18th European signal processing conference*. IEEE
77. Tolieng V et al (2017) Identification and lactic acid production of bacteria isolated from soils and tree barks. *Malays J Microbiol* 13(2):100–108
78. Khan R et al (2012) Color based skin classification. *Pattern Recognit Lett* 33(2):157–163
79. Shin MC, Chang KI, Tsap LV (2002) Does colorspace transformation make any difference on skin detection? In: *Proceedings of sixth IEEE workshop on applications of computer vision, 2002 (WACV 2002)*. IEEE
80. Schmugette SJ et al (2007) Task-based evaluation of skin detection for communication and perceptual interfaces. *J Vis Commun Image Represent* 18(6):487–495
81. Kasson JM, Plouffe W (1992) An analysis of selected computer interchange color spaces. *ACM Trans Gr (TOG)* 11(4):373–405
82. Xiong W, Li Q (2012) Chinese skin detection in different color spaces. In: *2012 international conference on wireless communications and signal processing (WCSP)*. IEEE
83. Araban S, Farokhi F, Kangarloo K (2011) Determining effective colour components for skin detection using a clustered neural network. In: *2011 IEEE international conference on signal and image processing applications (ICSIPA)*. IEEE
84. Beale MH, Hagan MT, Demuth HB (2010) *Neural network toolbox. User's Guide MathWorks* 2:77–81
85. Zolfaghari H, Nekonom AS, Haddadnia J (2011) Color-base skin detection using hybrid neural network & genetic algorithm for real times. *Int J Comput Sci Inf Secur* 9(10):67–71
86. Bhojar K, Kakde O (2010) Skin color detection model using neural networks and its performance evaluation. *J Comput Sci. Citeseer*
87. Doukim CA et al (2011) Combining neural networks for skin detection. *arXiv preprint arXiv:1101.0384*
88. Flach PA, Lachiche N (2004) Naive Bayesian classification of structured data. *Mach Learn* 57(3):233–269
89. Metzger A, Sammodi O, Pohl K (2013) Accurate proactive adaptation of service-oriented systems. In: *Cámara J, de Lemos R, Ghezzi C, Lopes C (eds) Assurances for self-adaptive systems*. Springer, Berlin, pp 240–265
90. Tsai C-W et al (2013) Data mining for internet of things: a survey. *IEEE Commun Surv Tutor* 16(1):77–97
91. Kant S, Ansari IA (2016) An improved K means clustering with Atkinson index to classify liver patient dataset. *Int J Syst Assur Eng Manag* 7(1):222–228
92. Erfani SM et al (2016) High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning. *Pattern Recognit* 58:121–134
93. Caterino N et al (2009) Comparative analysis of multi-criteria decision-making methods for seismic structural retrofitting. *Comput Aided Civ Infrastruct Eng* 24(6):432–445
94. Gayatri V, Chetan M (2013) Comparative study of different multicriteria decision-making methods. *Int J Adv Comput Theory Eng (IJACTE)* 2:9–12
95. Aruldoss M, Lakshmi TM, Venkatesan VP (2013) A survey on multi criteria decision making methods and its applications. *Am J Inf Syst* 1(1):31–43
96. Liu CH, Lin C-WR (2016) The comparative of the AHP TOPSIS analysis was applied for the commercialization military aircraft logistic maintenance establishment. *Int Bus Manag Spec* 4:6428–6432

97. Ashraf QM, Habaebi MH, Islam MR (2016) TOPSIS-based service arbitration for autonomic internet of things. *IEEE Access* 4:1313–1320
98. Singla C (2018) Modelling and analysis of multi-objective service selection scheme in IoT-cloud environment. In: Sangaiyah AK, Thangavelu A, Sundaram VM (eds) *Cognitive computing for big data systems over IoT*. Springer, Berlin, pp 63–77
99. Nunes LH et al (2017) Multi-criteria IoT resource discovery: a comparative analysis. *Softw Pract Exp* 47(10):1325–1341
100. Alsattar H et al. (2019) Novel meta-heuristic bald eagle search optimisation algorithm. *Artif Intell Rev* 52:21
101. Zaidan AA et al (2019) A new hybrid algorithm of simulated annealing and simplex downhill for solving multiple-objective aggregate production planning on fuzzy environment. *Neural Comput Appl* 31(6):1823–1834
102. Sameer FO et al (2019) A new algorithm of modified binary particle swarm optimization based on the Gustafson–Kessel for credit risk assessment. *Neural Comput Appl* 31(2):337–346
103. Abdullateef BN et al (2016) An evaluation and selection problems of OSS-LMS packages. *SpringerPlus* 5(1):248
104. Mansooreh M, Pet-Edwards J (1997) *Technical briefing: making multiple-objective decisions*. Institute of Electrical and Electronics Engineers Inc., IEEE Computer Society Press, Los Alamitos
105. Triantaphyllou E (2000) Multi-criteria decision making methods, in Multi-criteria decision making methods: a comparative study. Springer, Berlin, pp 5–21
106. Triantaphyllou E et al (1998) Multi-criteria decision making: an operations research approach. *Encycl Electr Electron Eng* 1998(15):175–186
107. Jumaah F et al (2018) Technique for order performance by similarity to ideal solution for solving complex situations in multi-criteria optimization of the tracking channels of GPS baseband telecommunication receivers. *Telecommun Syst*. <https://doi.org/10.1007/s11235-017-0401-5>
108. Kiah MLM et al (2014) Open source EMR software: profiling, insights and hands-on analysis. *Comput Methods Progr Biomed* 117(2):360–382
109. Qader M et al (2017) A methodology for football players selection problem based on multi-measurements criteria analysis. *Measurement* 111:38–50
110. Salman OH et al (2017) Novel methodology for triage and prioritizing using “big data” patients with chronic heart diseases through telemedicine environmental. *Int J Inf Technol Decis Mak* 16(05):1211–1245
111. Zaidan B et al (2017) A new approach based on multi-dimensional evaluation and benchmarking for data hiding techniques. *Int J Inf Technol Decis Mak*. <https://doi.org/10.1142/S0219622017500183>
112. Zaidan B et al (2017) A new digital watermarking evaluation and benchmarking methodology using an external group of evaluators and multi-criteria analysis based on ‘large-scale data’. *Softw Pract Exp* 47(10):1365–1392
113. Zaidan B, Zaidan A (2017) Software and hardware FPGA-based digital watermarking and steganography approaches: toward new methodology for evaluation and benchmarking using multi-criteria decision-making techniques. *J Circuits Syst Comput* 26(07):1750116
114. Zaidan B, Zaidan A (2018) Comparative study on the evaluation and benchmarking information hiding approaches based multi-measurement analysis using TOPSIS method with different normalisation, separation and context techniques. *Measurement* 117:277–294
115. Jumaah F et al (2018) Decision-making solution based multi-measurement design parameter for optimization of GPS receiver tracking channels in static and dynamic real-time positioning multipath environment. *Measurement* 118:83–95
116. Shih H-S, Shyur H-J, Lee ES (2007) An extension of TOPSIS for group decision making. *Math Comput Model* 45(7–8):801–813
117. Saaty TL, Ozdemir MS (2003) Why the magic number seven plus or minus two. *Math Comput Model* 38(3–4):233–244
118. Lesmes D, Castillo M, Zarama R (2009) Application of the analytic network process (ANP) to establish weights in order to re-accredit a program of a university. In: *Proceedings of the international symposium on the analytic hierarchy process*
119. Saaty TL (2008) Decision making with the analytic hierarchy process. *Int J Serv Sci* 1(1):83–98
120. Whaiduzzaman M et al (2014) Cloud service selection using multicriteria decision analysis. *Sci World J*. <https://doi.org/10.1155/2014/459375>
121. Çalışkan H (2013) Selection of boron based tribological hard coatings using multi-criteria decision making methods. *Mater Des* 50:742–749
122. Oztaysi B (2014) A decision model for information technology selection using AHP integrated TOPSIS-Grey: the case of content management systems. *Knowl Based Syst* 70:44–54
123. Albahri O et al (2018) Real-time remote health-monitoring systems in a medical centre: a review of the provision of healthcare services-based body sensor information, open challenges and methodological aspects. *J Med Syst* 42(9):164
124. Samvedi A, Jain V, Chan FT (2013) Quantifying risks in a supply chain through integration of fuzzy AHP and fuzzy TOPSIS. *Int J Prod Res* 51(8):2433–2442
125. Nilsson H, Nordström E-M, Öhman K (2016) Decision support for participatory forest planning using AHP and TOPSIS. *Forests* 7(5):100
126. Kalid N et al (2018) Based on real time remote health monitoring systems: a new approach for prioritization “large scales data” patients with chronic heart diseases using body sensors and communication technology. *J Med Syst* 42(4):69
127. Zaidan A et al (2015) Robust pornography classification solving the image size variation problem based on multi-agent learning. *J Circuits Syst Comput* 24(02):1550023
128. Taylan O, Kaya D, Demirbas A (2016) An integrated multi attribute decision model for energy efficiency processes in petrochemical industry applying fuzzy set theory. *Energy Convers Manag* 117:501–512
129. Barrios MAO et al (2016) An AHP-topsis integrated model for selecting the most appropriate tomography equipment. *Int J Inf Technol Decis Mak* 15(04):861–885
130. Albahri O et al (2019) Based multiple heterogeneous wearable sensors: a smart real-time health-monitoring structured for hospitals distributor. *IEEE Access* 7:37269–37323
131. Albahri A et al (2018) Real-time fault-tolerant mhealth system: comprehensive review of healthcare services, opens issues, challenges and methodological aspects. *J Med Syst* 42(8):137
132. Saaty TL (1990) How to make a decision: the analytic hierarchy process. *Eur J Oper Res* 48(1):9–26
133. Saaty TL (1977) A scaling method for priorities in hierarchical structures. *J Math Psychol* 15(3):234–281
134. Saaty TL, Vargas LG (1984) Inconsistency and rank preservation. *J Math Psychol* 28(2):205–214
135. Al-Azab FGM, Ayu MA (2010) Web based multi criteria decision making using AHP method. In: *Proceeding of the 3rd international conference on information and communication technology for the Moslem world (ICT4M) 2010*. IEEE
136. Rahmatullah B et al (2017) Multi-complex attributes analysis for optimum GPS baseband receiver tracking channels selection. In: *2017 4th international conference on control, decision and information technologies (CoDIT)*. IEEE

137. Salih MM et al (2018) Survey on fuzzy TOPSIS state-of-the-art between 2007–2017. *Comput Oper Res* 104:207–227
138. Alsalem M et al (2018) Systematic review of an automated multiclass detection and classification system for acute Leukaemia in terms of evaluation and benchmarking, open challenges, issues and methodological aspects. *J Med Syst* 42(11):204
139. Kalid N et al (2018) Based real time remote health monitoring systems: a review on patients prioritization and related” big data” using body sensors information and communication technology. *J Med Syst* 42(2):30
140. Chen C-T (2000) Extensions of the TOPSIS for group decision-making under fuzzy environment. *Fuzzy Sets Syst* 114(1):1–9
141. Huang Y-S et al (2013) Aggregation of utility-based individual preferences for group decision-making. *Eur J Oper Res* 229(2):462–469
142. Xia M, Chen J (2015) Multi-criteria group decision making based on bilateral agreements. *Eur J Oper Res* 240(3):756–764

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Affiliations

A. A. Zaidan<sup>1</sup> · B. B. Zaidan<sup>1</sup> · M. A. Alsalem<sup>2</sup> · O. S. Albahri<sup>1</sup> · A. S. Albahri<sup>3</sup> · M. Y. Qahtan<sup>1</sup>

<sup>1</sup> Department of Computing, Faculty of Arts, Computing and Creative Industry, Universiti Pendidikan Sultan Idris, Tanjung Malim, Malaysia

<sup>2</sup> College of Administration and Economic, University of Mosul, Mosul, Iraq

<sup>3</sup> Iraqi Commission for Computers & Informatics, Baghdad, Iraq