

# Skin cancer classification based on an optimized convolutional neural network and multicriteria decision-making

Neven Saleh

[nesaleh@msa.edu.eg](mailto:nesaleh@msa.edu.eg)

October University of Modern Sciences and Arts

Mohammed A Hassan

Helwan University

Ahmed M Salaheldin

El Shorouk Academy

---

## Article

**Keywords:** skin cancer, CNN, gray wolf optimizer, machine learning, multicriteria decision-making

**Posted Date:** March 14th, 2024

**DOI:** <https://doi.org/10.21203/rs.3.rs-3987239/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Additional Declarations:** No competing interests reported.

---

# Skin cancer classification based on an optimized convolutional neural network and multicriteria decision-making

Neven Saleh\*<sup>1,2</sup>, Mohammed A Hassan<sup>3</sup>, and Ahmed M Salaheldin<sup>1</sup>

<sup>1</sup> Systems and Biomedical Engineering Department, Higher Institute of Engineering, Shorouk Academy.

<sup>2</sup>Electrical Communication & Electronic Systems Engineering Department, Faculty of Engineering, October University for Modern Sciences and Arts (MSA).

<sup>3</sup> Biomedical Engineering Department, Faculty of Engineering, Helwan University.

\*Corresponding author e-mail: [nesaleh@msa.edu.eg](mailto:nesaleh@msa.edu.eg)

## Abstract:

Skin cancer can be treated if it is detected early. Many artificial intelligence-based models have been developed for skin cancer detection and classification. Considering the development of multiple models according to various scenarios and selecting the optimum model, these models were rarely considered in previous works. This study aimed to develop multiple models for skin cancer classification and select the optimum model. Convolutional neural networks (CNNs) in the form of AlexNet, Inception V3, MobileNet V2, and ResNet 50 were used for feature extraction. Feature reduction was carried out using two algorithms of the gray wolf optimizer (GWO) in addition to using the original features. Skin cancer images were classified into four classes based on six machine learning (ML) classifiers. As a result, 51 models were developed with different combinations of CNN algorithms, without GWO algorithms, with two GWO algorithms, and with six ML classifiers. To select the optimum model with the best results, the multicriteria decision-making approach was utilized in the recent form of ranking the alternatives by perimeter similarity (RAPS). Model training and testing were conducted using the International Skin Imaging Collaboration (ISIC) 2017 dataset. Based on nine evaluation metrics and according to the RAPS method, the AlexNet algorithm with GWO yielded the optimum model, achieving a classification accuracy of 94.5%. This work presents the first study on benchmarking skin cancer classification with a large number of models. Feature reduction not only reduces the time spent on training but also improves classification accuracy. The RAPS method has proven its robustness in the problem of selecting the best model for skin cancer classification.

**Keywords:** *skin cancer; CNN; gray wolf optimizer; machine learning; multicriteria decision-making.*

## 1. Introduction

A frequent type of cancer that is often caused by sun exposure is skin cancer [1]. Melanoma and nonmelanoma cases are classified according to the clinical category [1-3]. Even though melanoma is less common than nonmelanoma, it is the deadliest [2-5]. In 2019, the average death rate was 4740 males and 2490 females because of melanoma [1]. Indeed, the high cure rate of skin cancer mainly depends on early detection. In fact, dermatologists usually

depend on dermoscopic images for melanoma detection and diagnosis. Despite this being a common method for diagnosis, it is not accurate. This is due to the high similarity of features between melanoma and nonmelanoma lesions. Additionally, image noise in the form of blood vessels and hair [4] can lead to confusion. The accuracy of correctly diagnosing melanoma using visual examination is less than 80% [6].

To improve the diagnosis of melanoma, several automated systems have been proposed for identifying skin lesions to aid dermatologists. In this context, machine learning (ML) and deep learning methods were adopted. Various publications have classified melanoma and nonmelanoma lesions using ML methods. However, many limitations arise in terms of shallow training performance, extensive image processing, and variations in interclass and intraclass dermoscopic images [7]. Relevant studies were conducted based on ML methods. For instance, the K-nearest neighbor (K-NN) method has been employed for skin cancer classification, and it has achieved an accuracy of 98% [8].

Unlike in ML algorithms, large sets of raw data are used as inputs, and end-to-end training and automatic feature learning are the main advantages of deep learning algorithms [9]. In fact, the CNN technique has extensive applications in disease diagnosis. The performance of CNNs in a variety of cancer detection and classification tasks, such as breast cancer [10], prostate cancer [11], liver lesions [12], and lung cancer [13], is particularly encouraging. According to *Sudharshan et al.* [14], employing CNN models enhances the performance of diagnostic systems. The general CNN framework consists of layers for feature extraction, feature mapping, and subsampling [15]. However, feature extraction and mapping play a significant role in the outcome of the network. Taking this into account, the performance of some algorithms is influenced by the number of features, and rabbits in turn require feature reduction. The benefits of feature reduction include data storage space reduction, computation time reduction, removing noisy data, and improving the classification process [16].

Many techniques, including filtering, wrapping, and embedding, have been proposed for feature reduction. The heuristic search algorithm is a wrapper technique [16]. Metaheuristic methods, or stochastic methods, are adopted for tackling various problems as solvers when optimization is needed. These methods do not rely on the mathematical characteristics of the object [17]. Swarm intelligence-based stochastic approaches are the most widely used. The primary evolutionary origins of swarm-based approaches are drawn from animal foraging, hunting, and survival of the fittest [18]. Recent examples of well-known methods include the Harris hawk optimizer (HHO) [17], the whale optimization algorithm (WOA) [19], ant colony optimization (ACO) [20], fruit fly optimization (FFO) [21], and the gray wolf optimizer (GWO) [22].

This study was conducted to classify skin cancer using different types of CNNs. The main hypothesis of the study is whether optimizing the features will enhance the performance of the CNNs. The other question that arises is which is the best paradigm for classifying skin cancer. To answer these questions, we must select the best model according to the metrics of evaluation performance. In this context, the multicriteria decision-making (MCDM) approach is appropriate for selection. Therefore, the aim of this study was to select the best paradigm for skin cancer classification by optimizing the features and employing the MCDM approach. Thus, the authors employed three main approaches in this study: deep learning and ML, metaheuristic optimization, and MCDM. For deep learning, the AlexNet [23], Inception V3 [9], MobileNet V2 [24], and ResNet 50 [2] architectures were proposed. In addition, distinctive types of ML were employed. Among the available optimizers, the gray wolf optimizer was proposed [22]. For the MCDM stage, a recent method called ranking the alternatives by perimeter similarity (RAPS) [25] was investigated.

The main contributions of this study are summarized as follows: (i) different networks of CNNs with different combinations of hyperparameters are used to minimize biasing errors; (ii) features are optimized using a recent optimizer with different scenarios to improve classification accuracy; (iii) multiple ML classifiers with various configurations are used for image classification; (iv) for the first time, a significant number of models for the classification of skin cancer have been constructed in one study using different configurations; (v) a recent method of the MCDM approach was used to select the best CNN model; (vi) the importance of feature reduction in the classification of skin cancer and selection of the best model has been highlighted; and (vii) the automated diagnosis of skin cancer has improved, which in turn impacts diagnosis and patient healing.

The article is organized as follows: Related works on the problem at hand are presented in Section 2. The background of the utilized methods and tools is covered in Section 3. The methodology, which includes different approaches, is introduced in detail in Section 4. The experiments and results of each utilized approach associated with the discussion of the obtained results are demonstrated in Section 5. Finally, the study conclusions and future work are given in Section 6.

## **2. Related works**

Various studies have been presented on the problem of skin cancer detection and classification. Additionally, different CNN models were tested for this problem because most studies concluded that there was a significant enhancement in the classification accuracy of CNNs [26]. In 2020, *Zhang et al.* [27] presented a new methodology based on a CNN for detecting skin cancer. The method combines the WOA with a CNN to optimize its usage. Through integration, errors are minimized due to the optimum selection of weights and biases. The results demonstrated the efficiency of the study. *Singh et al.* [18]

developed a model to classify melanoma using inception and residual networks. The authors have added 40 layers for the classifiers. The models were tested using a dataset from the International Skin Imaging Collaboration (ISIC) for the years 2018, 2019, and 2020. The results demonstrated the effectiveness of the developed models compared to other algorithms. The performance of the models was evaluated through accuracy, specificity, sensitivity, and intersection over union (IOU) metrics.

In 2023, a set of dermoscopic images was split into melanoma and nonmelanoma images [2]. Two feature extraction and fine-tuning strategies have been used in transfer learning-based techniques. The classifiers EfficientNet B6, ResNet 50, DesNet 121, and Inception-ResNet V2 were tested for both techniques. The classifiers were evaluated through area under the curve (AUC) and receiver operating characteristic (ROC) curve analyses using datasets from ISIC 2019 and 2020. The results proved the robustness of EfficientNet B6 for both feature extractions, with an AUC of 0.9174, and fine tuning, with an AUC of 0.9681. In another relevant study, *Dong et al.* [3] proposed modifications to Inception-ResNet-V1 combined with quantum computing to improve the performance of CNNs. The support vector machine (SVM) method was employed in place of the network's fully connected (FC) layer. Additionally, to lessen the effects of data imbalance, the authors used weighted sampling and data augmentation techniques. This study adopted three experiments for the ISIC 2019 dataset to test the proposed approach. The best performance was measured for the experiment using a processed dataset, for which the accuracy was 98.76%. An improved artificial rabbit optimizer was introduced to optimize features for detecting skin cancer. The MobileNetV3 network has been employed to detect melanoma-related skin cancer. Three datasets were tested to prove the model's validity. The model yielded rational results with accuracies of 87.17%, 96.79%, and 88.71% for the different datasets [4].

In another relevant work, a stacked convolutional neural network was proposed for detecting melanoma [5]. First, the authors used pretrained CNN models to demonstrate the performance of the modified CNN model. Second, the stacked CNN, which uses 2D layers, was presented. The improved network was evaluated for two datasets, MINST HAM1000 and ISIC 2020. A t test was carried out to determine the significance of the proposed network compared to traditional networks. A new framework for skin cancer detection was presented using MobileNetV3 with a novel algorithm for feature selection [28]. Once the features have been extracted, a modified Hunger Games Search (HGS) based on Particle Swarm Optimization (PSO) and Dynamic-Opposite Learning (DOLHGS) was utilized as the input. The proposed paradigm was tested on the ISIC 2016 and PH2 datasets. Multiple classification models based on deep learning were presented in [29]. In this study, a novel CNN was presented for classifying skin cancer. The network is called a deep learning-based skin cancer classification network (DSCC\_Net). The network was compared to six well-known CNN models: ResNet-152, VGG-19, VGG-16, EfficientNet-B0, MobileNet,

and Inception-V3. DSCC\_Net was tested using three different datasets (ISIC 2020, DermIS, and HAM10000). The network has yielded good performance compared to the other networks.

Another study classified skin cancer based on training various deep learning-based methods. Two datasets were used for validation: ISIC 2019 and ISIC 2020 [30]. In 2022, *Javed Rashid* et al. [24] published a study on detecting skin cancer using a transfer learning technique. The authors proposed a novel transfer learning CNN using MobileNet V2. The ISIC 2020 dataset was used to test the model. Classification of different types of skin cancer was introduced by *Naeem* et al. [31], in which a novel CNN called SCDNet was proposed based on VGG16. Different CNN models, such as ResNet50 and VGG19, were compared against the developed model. The dataset from ISIC 2019 was used for comparison and validation of the network. Thus, CNNs have been employed in different versions for skin cancer detection. This has encouraged us to employ various CNN models. The findings of feature selection optimization are also promising. Additionally, we noted that the MCDM approach was not adopted for selecting the best model out of the various models. In this way, the authors decided to combine these approaches to develop an improved automated system for skin cancer classification.

### **3. Background**

This work adopted three approaches for implementation. First, a deep learning approach is used to construct CNNs with different structures. Second, an optimization is employed for feature reduction and selection. Third, MCDM was used to select the best paradigm for classifying skin cancer. A general overview of each approach is given in the following subsections. The ML classifiers utilized in the classification process were introduced in the Materials and Methods section.

#### **3.1 Convolution Neural Network (CNN)**

The CNN is the most popular deep learning algorithm. The origin of the network is referred to as the appearance of LeNet in 1989 [32, 33]. The main advantage of CNNs is their ability to extract features automatically without guiding supervision. In addition, compared to other neural networks, CNNs are significantly easier to construct on a large scale [32]. In application, CNNs have widespread applications in various fields, such as disease detection and classification [2, 9, 10, 11, 34], transportation [35], facial recognition [36], and speech recognition [37]. A common architecture structure of CNNs is convolutional layers followed by subsampling (pooling) layers followed by FC layers [32]. Many modified CNN layouts have been created to improve the performance of CNNs. Examples of modified CNNs include AlexNet, ResNet 50, Inception V3/V4, GoogleNet, MobileNet-v2, VGG 16/19, SqueezeNet, EfficientNet-B0, and HRNetV2 [9, 29, 32]. The study employs AlexNet [23], Inception V3 [9], MobileNet V2 [24], and ResNet 50 [38]. Therefore, a brief description of each network is given in the following subsections.

### 3.1.1 AlexNet

AlexNet was created by *Krizhevsky* et al. [39] in 2012. To enhance the learning capability, many parameters have been optimized in addition to the increase in length. The depth of the AlexNet is 8 layers, and the basic input image size is  $227 \times 227 \times 3$ . Several origin-related drawbacks of AlexNet include the extension of the five phases of feature extraction to seven stages [32]. In AlexNet, ReLU is used as a nonsaturating activation function. Moreover, large filters  $5 \times 5$  and  $11 \times 11$  are employed by the AlexNet [32, 40].

### 3.1.2 Inception-v3

Inception V3 was designed depending upon its previous version, Inception V1/2. Improvements include reducing the computational cost by selecting small filter sizes of  $1 \times 5$  and  $1 \times 7$ . In addition, the proposed method utilizes a small convolution layer ( $1 \times 1$ ) before large filters are used. As a result, compared with that of Inception V2, the performance of the model is reduced by 1.38 times [41]. However, the network comprises approximately 20 million parameters [42]. The soft-max layer is used for classification, and batch normalization is used for the activation layer [42]. Generally, the inception V3 is composed of 48 layers with an image size of  $229 \times 229 \times 3$  [9, 32].

### 3.1.3 ResNet 50

The residual neural network (ResNet 50) was introduced by *He* et al. [26] in 2016. The idea behind this network is to improve the performance by adding residual connections between layers. As a result, losses decrease, knowledge gain increases, training performance improves, and the model becomes robust against overfitting [42]. Several layers compose the ResNet model. It starts with 34 layers and ends with 1202 layers. This variation leads to changes in the residual block number and basic operations. The most common type is ResNet 50, which is composed of 49 convolutional layers and one FC layer [43].

### 3.1.4 MobileNet V2

One form of CNN that works for optimizing mobile devices is MobileNet V2. The main advantage of this network is its ability to classify images with less computational effort [44]. In this way, a supercomputer is not mandatory. The architecture of the network comprises depthwise and pointwise convolution layers [45]. The network was developed based on MobileNet V1. However, MobileNet V2 is more reliable than MobileNet V1. This is due to two factors: linear bottlenecks between the layers and shortcut connections between the bottlenecks [44, 45].

### 3.2 Gray Wolf Optimizer

Swarm intelligence (SI) is a powerful form of computational intelligence inspired by the behavior of a natural swarm [22]. The main idea of SI-based algorithms is to simulate the behavior of swarm members when finding food or hunting prey [46]. This principle is applied by numerous algorithms to optimize a solution to a problem. Examples of SI algorithms include the ant colony algorithm (ACO) [47], the WOA [27], the artificial bee colony (ABC) [48], and the gray wolf optimizer (GWO) [22, 49]. In 2014, *Mirjalili et al.* [50] introduced the GWO, a new SI-based algorithm. In nature, wolves are hunting prey optimally. The algorithm applies this principle by organizing wolves' roles in the pack according to the hierarchy [50, 51]. Based on the many roles that wolves play in the hunting process, the members of the pack in GWO are separated into four groups. The distinctive groups are alpha, beta, delta, and omega, with alpha denoting the greatest hunting solution [49-51]. The wolf's category beta assists in hunting decisions, while the third-class delta surrounds the first and second classes for scouting and hunting. Finally, the omega wolves play the lowest role by protecting the backs of the previous wolves. Thus, the roles of wolves in hunting prey can be identified as tracking, pursuing, and assaulting [49]. The mathematical modeling of the GWO according to the social behavior of the wolves is explained in the following steps [22, 49-51].

Assume that the initial location of a wolf to a prey in a circle is provided by the distance  $D$ , as shown in Equation 1. Due to the chasing process, the location of the wolf is updated through Equation 2.

$$D = |C \cdot X_p(t) - X(t)| \quad (1)$$

$$X(t+1) = X_p(t) - A \cdot D \quad (2)$$

where  $D$  is the distance between the wolf and the prey,  $t$  is the number of iterations,  $X(t)$  is the current location of the wolf,  $X_p(t)$  is the location of the prey, and  $X(t+1)$  is the next location of the wolf. The coefficients  $A$  and  $C$  are calculated by Equations 3 and 4, respectively.

$$A = (2 \times a \times r_1) - a \quad (3)$$

$$C = 2 \times r_2 \quad (4)$$

To improve the movement of the wolf, vectors  $r_1$  and  $r_2$  are randomly selected between 0 and 1. The coefficient  $a$  provides a linear decay range from 2 to 0, as presented in Equation 5, where  $t$  is the number of iterations and  $T$  is the maximum number of iterations.

$$a = 2 - \left(\frac{2t}{T}\right) \quad (5)$$

According to the previous equations and with numerous iterations, the locations of all wolves (alpha, beta, delta, and omega) are determined and updated toward the prey's location using Equation 6.

$$\left. \begin{aligned} D_\alpha &= |C_1 X_\alpha - X| \\ D_\beta &= |C_2 X_\beta - X| \\ D_\delta &= |C_3 X_\delta - X| \end{aligned} \right\} \quad (6)$$

$$\left. \begin{aligned} X_1 &= |X_\alpha - A_1 \cdot D_\alpha| \\ X_2 &= |X_\beta - A_1 \cdot D_\beta| \\ X_3 &= |X_\Omega - A_1 \cdot D_\Omega| \end{aligned} \right\} \quad (7)$$

where a wolf alpha and a wolf omega are separated by  $D_\alpha$ , a wolf beta and a wolf omega are separated by a distance  $D_\beta$ , and a wolf delta and a wolf omega are separated by a distance  $D_\delta$ . The new optimum locations for the alpha, beta, and delta wolves are defined as  $X_1$ ,  $X_2$ , and  $X_3$ , respectively, as illustrated in Equation 7. By assuming that the best solutions for each category are found, other wolf locations are determined, as shown in Equation 8.

$$X(t+1) = (X_1 + X_2 + X_3)/3 \quad (8)$$

### 3.3 Multi-Criteria Decision-Making Approach

The problem of selecting an alternative among numerous options requires a comprehensive solution. To resolve this problem, many aspects should be considered, including the criteria of selection and their weights. The MCDM approach tackles such problems with many techniques to promote nonbiasing decision-making. The beginning of the MCDM was in the early eighteenth century [25]. Modern common techniques were first introduced in 1965 by developing elimination and choice in expressing reality (ELECTRE). Several modifications were added to ELECTRE I to produce ELECTRE II, ELECTRE III, ELECTRE IV, and ELECTRE as separate models [25]. In 1968, MacCrimmon created a different technique known as simple additive weighting (SAW) [52]. In the 1970s, Saaty developed the analytic hierarchy process (AHP) as a tool for MCDM [53, 54]. Since then, numerous methods have been introduced, including but not limited to the technique for

order preference by similarity to the ideal solution (TOPSIS), multiobjective optimization by ratio analysis (MOORA), the preference ranking organization method for enrichment evaluation (PROMETHEE), data envelopment analysis (EDA), case-based reasoning (CBR), and the additive ratio assessment system (ARAS) [53, 54].

As the degree to which a criterion is preferred obviously affects the decision-making process, assessing the weight of the criterion is an essential stage in the MCDM. Objective-based methods are widely used for criteria weighting. The entropy method and criterion importance via intercriteria correlation (CRITIC) are common examples of this method [53]. To implement a standard MCDM method, one must first identify the problem, then create acceptable criteria, weigh the criteria, identify potential solutions, and use mathematical formulas to apply a chosen MCDM technique [55]. One of the most recent MCDM techniques is called ranking alternatives by perimeter similarity (RAPS) [25]. In 2021, Urošević et al. [56] developed the RAPS method for its first application in the mining industry. The application of the RAPS method is explained in the following steps [25, 55, 56]:

**Step 1:** Normalize the initial data that describe the alternatives against the criteria. Considering that the selected criteria are either beneficial, i.e., maximization is needed, or nonbeneficial, i.e., minimization is required [53]. Equations 9 and 10 present normalizations for both beneficial and nonbeneficial criteria, respectively.

$$r_{ij} = x_{ij}/Max_i(x_{ij}), \quad i=1, \dots, n \quad \& \quad j=1, \dots, m \quad (\text{beneficial}) \quad (9)$$

$$r_{ij} = Min_i(x_{ij})/x_{ij} \quad i=1, \dots, n \quad \& \quad j=1, \dots, m \quad (\text{nonbeneficial}) \quad (10)$$

where  $r$  represents the normalized value of input  $x$  for an alternative  $i$  for  $n$  alternatives against a criterion  $j$  for  $m$  criteria.

**Step 2:** Calculate a weighted normalized matrix  $u_{ij}$ , where  $w_j$  represents each criterion weight, as shown in Equation 11.

$$u_{ij} = w_j \times r_{ij} \quad (11)$$

**Step 3:** Eligate each element of the best solution ( $q_j$ ) using Equation 12, which in turn produces an optimal solution ( $Q$ ) through Equation 13.

$$q_j = \max u_{ij}, \quad i=1, \dots, n \quad \& \quad j=1, \dots, m \quad (12)$$

$$Q = \{q_1, q_2, q_3, \dots, q_m\} \quad (13)$$

**Step 4:** Decomposition of an optimal alternative into two subcomponents  $Q^{max}$  and  $Q^{min}$  is performed. The maximum number of beneficial criteria is denoted by  $k$  to produce  $Q^{max}$ ; meanwhile, the maximum number of nonbeneficial criteria is denoted by  $h = m - k$  to yield  $Q^{min}$ . The vector  $Q$  is the union of the two subcomponents, as indicated by Eq. 14.

$$Q = Q^{max} \cup Q^{min} = \{q_1, q_2, \dots, q_k\} \cup \{q_1, q_2, \dots, q_h\}; k+h = m \quad (14)$$

**Step 5:** As in the previous step, each alternative is broken down into subsets  $U^{max}$  and  $U^{min}$  to produce  $U$  values, as shown in Equation 15.

$$U_i = U_i^{max} \cup U_i^{min} = \{u_{i1}, u_{i2}, \dots, u_{ik}\} \cup \{u_{i1}, u_{i2}, \dots, u_{ih}\}; i=1, \dots, n \quad (15)$$

**Step 6:** This step pertains to the component's magnitude and requires the calculation of every element that makes up the best option, as shown in Equations (16) and (17). The magnitude of each component is determined using Equations (18) and (19).

$$Q_k = \sqrt{q_1^2 + q_2^2 + \dots + q_k^2} \quad (16)$$

$$Q_h = \sqrt{q_1^2 + q_2^2 + \dots + q_h^2} \quad (17)$$

$$U_{ik} = \sqrt{u_{i1}^2 + u_{i2}^2 + \dots + u_{ik}^2} \quad ; i=1, \dots, n \quad (18)$$

$$U_{ih} = \sqrt{u_{i1}^2 + u_{i2}^2 + \dots + u_{ih}^2} \quad ; i=1, \dots, n \quad (19)$$

**Step 7:** Final sorting of alternatives is achieved by applying Equations 20-22. In terms of alternatives, the right-angle triangle perimeter  $P$  is the best option. Equation 20 gives the expression for components  $Q_k$  and  $Q_h$ , which correspond to the base and perpendicular sides of this triangle, respectively. For each alternative, Equation 21 is used to calculate an alternative perimeter,  $P_i$ . The ratio between  $P_i$  and  $P$  is used to determine an alternative ranking index  $PS_i$ , as demonstrated in Equation 22.

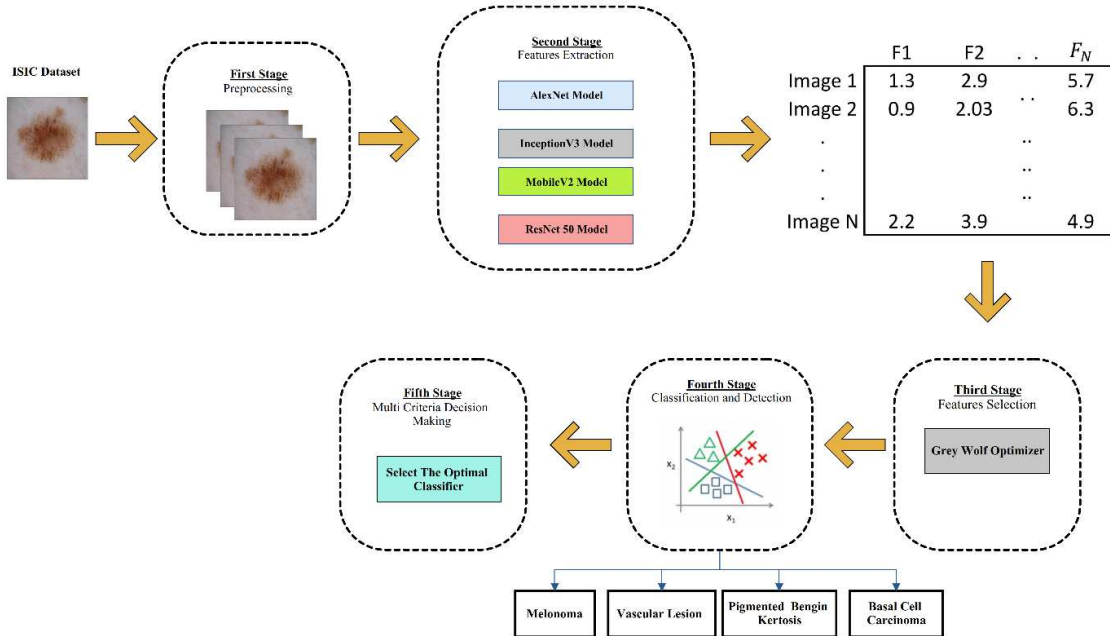
$$P = Q_k + Q_h + \sqrt{Q_k^2 + Q_h^2} \quad (20)$$

$$P_i = U_{ik} + U_{ih} + \sqrt{U_{ik}^2 + U_{ih}^2} \quad (21)$$

$$PS_i = P/P_i \quad ; i=1, \dots, n \quad (22)$$

## 4. Materials and Methods

This study presents a framework for classifying skin cancer. Various approaches have been applied to the present work. The CNN has four algorithms, the WGO for feature optimization, different ML classifiers, and the MCDM for selecting the optimum model. An overview of the proposed methodology is given in Fig. 1. As noted, the methodology was conducted in five stages, including distinctive methods. The first stage involved data processing; the second stage involved feature extraction via the CNN algorithm; the third stage involved feature selection via the application of GWO; the fourth stage involved classifying the dermoscopic images into four classes via different ML classifiers; and the fifth stage involved selecting the optimum model among all the developed models via the MCDM.

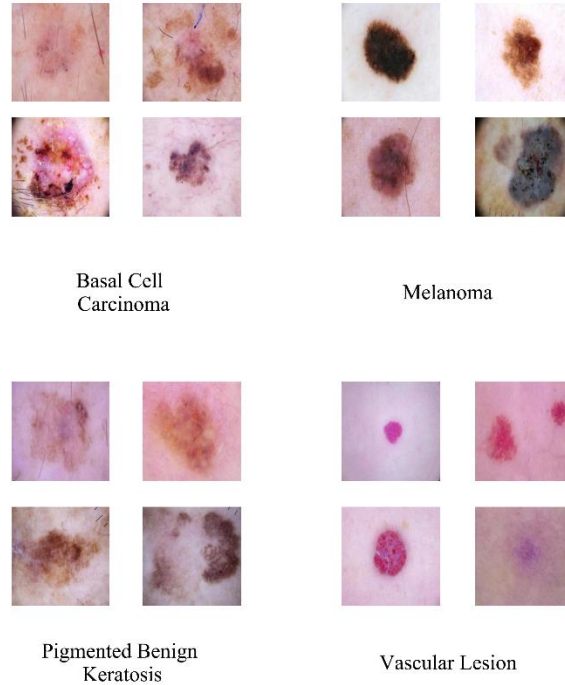


**Fig. 1** The proposed methodology for skin cancer classification in five stages, including distinctive methods.

### 4.1 Dataset Description

The employed dataset was obtained from the esteemed International Skin Imaging Collaboration (ISIC), specifically from its 2017 release [57]. This rich dataset encompasses a diverse array of nine distinct classes, each representing a unique dermatological condition. The classes were actinic keratosis, basal cell carcinoma, dermatofibroma, melanoma, nevus, pigmented benign keratosis, seborrheic keratosis, squamous cell carcinoma, and vascular lesions. A total of 2357 images were meticulously distributed across these classes, creating a comprehensive repository of dermoscopic visual data. In the process of feature extraction, certain classes are excluded due to their limited

representation compared to the remaining classes. The refined subset of classes comprises basal cell carcinoma, melanoma, pigmented benign keratosis, and vascular lesions, amounting to a total of 1466 images. Figure 2 shows examples of the investigated skin cancer classes. To optimize the model's efficacy, the dataset was judiciously partitioned into distinct sets: 70% for training, 15% for validation, and 15% for testing purposes.



**Fig. 2** Examples of the investigated skin cancer classes: basal cell carcinoma, melanoma, pigmented benign keratosis, and vascular lesions according to the ISIC 2017.

## 4.2 Data processing

The initial phase in crafting the envisioned model involves managing the dataset by adjusting its dimensions to align seamlessly with the input size requirements of each employed deep learning model. Remarkably, the dataset was of commendable quality, rendering any augmentation or additional image processing enhancement strategies unnecessary. Its inherent excellence obviates the need for further refinement, affirming its suitability for direct integration into model development.

## 4.3 Feature extraction for skin cancer images

When embarking upon the deployment of a computer-aided diagnosis (CAD) system, one encounters a pivotal decision regarding the extraction of features from the input data. Two fundamental approaches have been adopted to address this challenge. The first approach revolves around crafting features manually, utilizing traditional, handcrafted techniques. This method, though meticulously executed, often results in a limited number of features that may not faithfully capture the intricacies of the input data. This approach, while precise, can sometimes fall short in representing the full spectrum of characteristics

inherent to the data. The second involves harnessing the power of automated feature extraction through cutting-edge deep learning CNN algorithms. Leveraging CNNs unlocks the potential for generating an extensive array of features, offering a more comprehensive representation of each class or category within the dataset.

In our specific scenario, we have meticulously crafted and fine-tuned four distinct CNN models: AlexNet [23, 40], InceptionV3 [9], MobileNet V2 [24, 58], and ResNet-50 [2,43]. These models served as feature extractors for the representation of the various skin cancer classes under investigation. The quantity of extracted features is contingent upon the specific layer selected for feature extraction, effectively resulting in an output array denoted as  $\{F_1 F_2 F_3 \dots \dots F_N\}$ . A comprehensive overview of the feature extraction phase is provided in Table 1. The table provides a summary of the four scenarios, encompassing the employed CNN model, the total count of extracted features, and the layer from which these features were derived. This approach enables us to precisely capture and differentiate the distinctive characteristics of the targeted skin cancer classes, enhancing the efficacy of the CAD system. In the feature extraction phase, the training parameters were configured for a comprehensive span of 30 epochs, employing an initial learning rate set at 0.0001 and following a piecewise learning rate schedule. Validation through a hands-on approach was conducted, and the weights were updated after every 600 iterations. Striking a balance between efficiency and accuracy, a judicious mini-batch size of 64 was selected for the optimization process.

**Table 1** A summary of the feature extraction stages for skin cancer detection and classification.

<b>Model</b>	<b>Number of Features</b>	<b>Layer of Extraction</b>
AlexNet	4096	Fully Connected no. 6
InceptionV3	2048	Average Pooling
MobileNet V2	1280	Global Average Pooling
ResNet-50	2048	Global Average Pooling

#### 4.4 Feature selection for skin cancer images

The gray wolf optimizer (GWO) [49-51] algorithm was employed to aid in feature extraction. By applying different deep learning models during the development of a CAD system, four distinct categories of skin cancer were identified. Inspired by the hunting behavior of gray wolves, the GWO algorithm was used to fine-tune the feature selection process, ensuring the selection of the most informative and discriminative features for the classification task. Several parameters govern the GWO algorithm, including population size, iteration count, exploration and exploitation rates, and convergence threshold [50]. These parameters play a pivotal role in guiding the search for an optimal subset of features, enhancing the performance of the CAD system in classifying skin cancer. These values are proposed based on the assumption that GWO2 involves a more aggressive reduction in the number of features than does GWO1. The convergence threshold has also been adjusted to reflect the potentially more challenging optimization problem in GWO2.

During this stage, a modification was performed to the classical GWO, particularly for coefficient  $a$ , as denoted by Equation 23, to increase the exploration time for finding more solutions in the search space [51]. This algorithm was termed GWO1, while the algorithm related to the classical GWO was termed GWO2.

$$a = 2 - \left(\frac{2t^2}{T^2}\right) \quad (23)$$

The two versions were employed to extract two feature maps for each deep learning model based on the control parameter settings, as presented in Table 2. As shown in Table 2, the exploration and exploitation rates, in addition to the convergence threshold, were altered between the two versions. The optimization process reduces the number of features retained as the most significant for the classification phase. According to the flow of the GWO, as shown in Fig. 3, the fitness function was calculated for each wolf for each iteration. The proposed fitness function for both algorithms is presented in Equation 24.

$$F_{\text{GW}} = \frac{1}{T_{\text{max}}} \sum_{t=1}^{T_{\text{max}}} |N(t)_{\text{ref}}^{\text{label}} - N(t)_{\text{non}}^{\text{label}}| \quad (24)$$

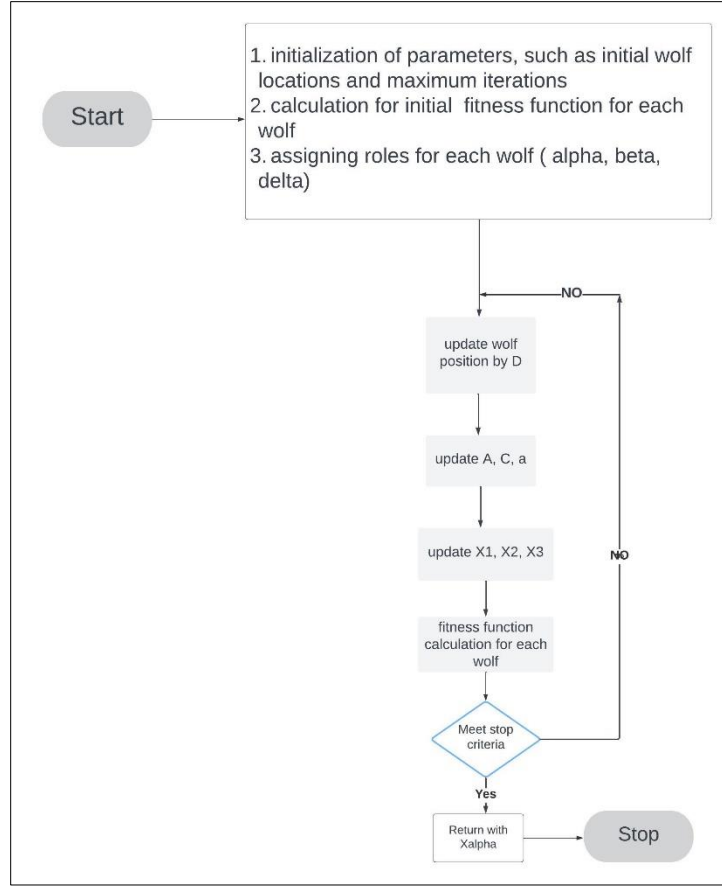
where  $F$  is the fitness function,  $N_{\text{ref}}$  is the output of the corresponding features for each labeled dermoscopic image, and  $N_{\text{non}}$  is the output of the noncorresponding features for each labeled dermoscopic image.

**Table 2** Controlling parameter settings for the two GWO algorithms.

Parameter	GWO 1	GWO 2
Population Size	10	10
Number of Iterations	100	100
Exploration Rate	2	1.5
Exploitation Rate	0.5	0.3
Convergence Threshold	$1 \times e^{-6}$	$4 \times e^{-7}$

#### 4.5 Machine learning classification for skin cancer images

The fourth pivotal stage within the outlined framework entails the execution of image classification. This involves training an array of diverse ML classifiers, spanning from the linear SVM [59] to the cubic SVM [43] and further extending to the quadratic SVM [58]. Additionally, model development embraces the incorporation of a medium-sized neural network [60], a wide neural network [61], and an ensemble subspace discriminant [62]. This comprehensive approach reflects a nuanced strategy, leveraging an eclectic mix of classifiers to intricately capture and discern patterns within the dataset, fostering a robust and versatile classification model for CAD systems.



**Fig. 3** Flowchart of both versions of GWO for feature reduction in skin cancer images.

#### 4.1 Application of the ARAS method for selecting the optimum model

The final stage of the proposed methodology involves selecting the optimum model for image classification. According to the sequence of our methodology, we have four feature extractors and three feature maps: the original features and two maps due to the two versions of the GWO. All combinations of the feature extractor and feature map were applied to all the ML classifiers. As a result, 51 models were developed for the classification of skin cancer. In this way, we need to apply the ARAS method to select the optimum model. A detailed description of the ARAS method is given in Section 3.3. According to the principles of the MCDM, a set of criteria was selected for model benchmarking. For this reason, nine evaluation metrics were chosen to distinguish the models. These included the accuracy (ACC), sensitivity (SV), specificity (SP), precision (PR), error rate (ER), false positive rate (FPR), false negative rate (FNR), negative predictive value (NPV), and F1 score (F1S) [9, 59]. All criteria are identified in terms of true positives (TPs), false positives (FPs), true negatives (TNs), and false negatives (FNs), as shown in Equations 25-33.

$$Accuracy = (TP + TN)/(TP + FN + FP + TN) \quad (25)$$

$$\text{Sensitivity} = TP/(TP + FN) \quad (26)$$

$$\text{Error rate} = (FP + FN)/(TP + FN + FP + TN) \quad (27)$$

$$\text{Specificity} = TN/(TN + FP) \quad (28)$$

$$\text{Precision} = TP/(TP + FP) \quad (29)$$

$$\text{False Positive Rate} = FP/(FP + TN) \quad (30)$$

$$\text{False Negative Rate} = FN/(FN + TP) \quad (31)$$

$$\text{Negative Predictive Value} = TN/(TN + FN) \quad (32)$$

$$F1\text{-Score} = (2 \times (\text{Sensitivity} \times \text{Precision})) / (\text{Sensitivity} + \text{Precision}) \quad (33)$$

Obviously, criteria weighting is a pivotal step in implementing the ARAS method. According to previous studies, entropy and CRITIC are the dominant methods. Two attempts were made to use the CRITIC method in this study. The procedure of the CRITIC method is explained through the next steps [53].

**Step 1:** Use the maximum and minimum values for each criterion, as indicated in Equation 34, to normalize each  $x_{ij}$  in the initial decision matrix to produce a normalized value  $r_{ij}$ .

$$r_{ij} = \frac{x_{ij} - x_j^{\min}}{x_j^{\max} - x_j^{\min}} \quad (34)$$

**Step 2:** A correlation coefficient is determined for all criteria as presented in (35). where  $\sigma$  is the standard deviation for each criterion  $j$ .

$$c_j = \sigma_j \sum_{i=1}^m (1 - r_{ij}) \quad (35)$$

**Step 3:** By normalizing the correlation coefficient, the criterion weight is computed as shown in Equation 35.

$$w_j = \frac{c_j}{\sum_{i=1}^m c_i} \quad (36)$$

## 5 Results and Discussion

This study was conducted to classify four classes of skin cancer. A total of 1466 images related to the ISIC 2017 were used to classify basal cell carcinoma, melanoma, pigmented benign keratosis, and vascular lesions. A distinctive framework was developed by applying deep learning algorithms, two versions of an optimization algorithm, ML classifiers, and the RAPS method to the dataset. Three scenarios were adopted for feature extraction: the

use of the original feature map, an optimized feature map due to the modified GWO called GWO1 and the classical optimizer known as GWO2. The numbers of selected features for each model for both algorithms are detailed in Table 3.

**Table 3** Number of selected features for the two GWO algorithms compared to the original features.

Model	GWO1	GWO2	Original Features
AlexNet	2248	1478	4096
InceptionV3	1476	683	2048
MobileNet V2	923	333	1280
ResNet-50	1420	776	2048

As noted in Table 3, for the AlexNet model, the number of features was reduced by 45% and 64% for GWO1 and GWO2, respectively. For the second model, the feature reductions were 28% and 67%, respectively. Considering the third model, the values were 28% and 74% for GWO1 and GWO2, respectively. Finally, for the ResNet-50 model, 31% and 64% of the samples were recorded for GWO1 and GWO2, respectively. These percentages revealed the performance of each algorithm separately. Furthermore, these findings clarify how strongly GWO2 affects feature selection.

Within each of these scenarios, we meticulously trained six ML classifiers to regulate the classification process within the tailored CAD system. For AlexNet (A), Inception V3 (I), and ResNet 50 (R), we trained a cubic SVM (CSVM), wide neural network (WNN), quadratic SVM (QSVM), and medium neural network (MNN) separately for each scenario. Furthermore, for MobileNet V2 (M), we trained the CSVM, WNN, QSVM, linear SVM (LSVM), and ensemble subspace discriminant (ESD) models separately for each scenario. As a result, 51 models were developed for classifying skin cancer, as shown in Table 4. All the models were compared against the evaluation criteria. The computer used to conduct this study included an Intel Core i7, an NVIDIA GeForce MX 130, Windows 11, 64 bits, and 16 GB of RAM. Moreover, all the experiments were performed using the MATLAB R2021b program.

**Table 4** Classification results of the developed models for skin cancer according to the three feature maps

Fet. maps	No	Model	ACC	SV	SP	PR	ER	FPR	FNR	NPV	FIS
GWO1	1	A.CSVM	0.945	0.883	0.961	0.892	0.054	0.038	0.116	0.962	0.887
	2	A.WNN	0.945	0.899	0.961	0.908	0.054	0.039	0.100	0.961	0.903
	3	A.QSVM	0.942	0.877	0.958	0.897	0.057	0.041	0.122	0.959	0.885
	4	A.MNN	0.941	0.892	0.958	0.894	0.058	0.041	0.107	0.958	0.893
GWO2	5	A.CSVM	0.943	0.880	0.959	0.900	0.056	0.040	0.119	0.960	0.888
	6	A.WNN	0.941	0.887	0.958	0.898	0.058	0.042	0.112	0.958	0.892
	7	A.QSVM	0.938	0.872	0.956	0.890	0.061	0.044	0.127	0.956	0.879

	8	A.MNN	0.939	0.878	0.956	0.897	0.060	0.043	0.122	0.957	0.887
<b>Orig.</b>	9	A.CSV M	0.945	0.883	0.960	0.903	0.054	0.039	0.116	0.961	0.892
	10	A.WNN	0.943	0.892	0.959	0.900	0.057	0.040	0.108	0.959	0.895
	11	A.QSV M	0.940	0.875	0.957	0.893	0.059	0.042	0.124	0.958	0.883
	12	A.MNN	0.940	0.885	0.957	0.895	0.059	0.042	0.114	0.957	0.890
<b>GW O1</b>	13	I. CSVM	0.925	0.826	0.947	0.855	0.074	0.052	0.173	0.949	0.837
	14	I.WNN	0.921	0.834	0.945	0.841	0.078	0.055	0.165	0.945	0.837
	15	I.QSVM	0.920	0.821	0.943	0.840	0.079	0.056	0.178	0.944	0.829
	16	I.MNN	0.920	0.838	0.944	0.832	0.079	0.055	0.161	0.944	0.835
<b>GW O2</b>	17	I. CSVM	0.915	0.809	0.940	0.836	0.084	0.060	0.190	0.941	0.820
	18	I.WNN	0.916	0.826	0.941	0.825	0.083	0.058	0.173	0.941	0.826
	19	I.QSVM	0.913	0.806	0.938	0.835	0.086	0.061	0.193	0.939	0.817
	20	I.MNN	0.884	0.703	0.914	0.702	0.115	0.085	0.296	0.914	0.702
<b>Orig.</b>	21	I. CSVM	0.929	0.840	0.950	0.862	0.070	0.049	0.159	0.951	0.850
	22	I.WNN	0.928	0.846	0.950	0.852	0.071	0.050	0.153	0.950	0.849
	23	I.QSVM	0.929	0.846	0.950	0.861	0.070	0.049	0.153	0.951	0.852
	24	I.MNN	0.925	0.841	0.948	0.847	0.074	0.052	0.158	0.948	0.844
<b>GW O1</b>	25	M.CSV M	0.885	0.764	0.917	0.804	0.115	0.082	0.235	0.919	0.782
	26	M.WNN	0.886	0.776	0.919	0.781	0.114	0.080	0.223	0.919	0.778
	27	M.QSV M	0.883	0.764	0.916	0.795	0.116	0.083	0.235	0.917	0.778
	28	M.LSV M	0.886	0.765	0.918	0.802	0.114	0.081	0.235	0.919	0.781
	29	M.SD	0.861	0.727	0.902	0.733	0.138	0.097	0.272	0.902	0.730
<b>GW O2</b>	30	M.CSV M	0.869	0.730	0.907	0.765	0.130	0.093	0.269	0.908	0.745
	31	M.WNN	0.864	0.735	0.904	0.740	0.136	0.095	0.264	0.904	0.737
	32	M.QSV M	0.869	0.732	0.907	0.762	0.130	0.092	0.267	0.908	0.745
	33	M.LSV M	0.874	0.732	0.910	0.787	0.125	0.090	0.267	0.911	0.755
	34	M.SD	0.869	0.748	0.907	0.760	0.130	0.092	0.251	0.908	0.754
<b>Orig.</b>	35	M.CSV M	0.882	0.760	0.915	0.799	0.117	0.084	0.239	0.917	0.777
	36	M.WNN	0.882	0.774	0.916	0.779	0.117	0.083	0.225	0.916	0.776
	37	M.QSV M	0.882	0.762	0.916	0.795	0.117	0.084	0.237	0.916	0.777
	38	M.LSV M	0.885	0.759	0.918	0.801	0.114	0.081	0.240	0.919	0.777
	39	M.SD	0.834	0.686	0.883	0.672	0.165	0.116	0.313	0.883	0.678
<b>GW O1</b>	40	R. CSVM	0.929	0.844	0.949	0.864	0.070	0.050	0.155	0.951	0.851
	41	R.WNN	0.931	0.850	0.950	0.870	0.068	0.049	0.149	0.952	0.858

	42	R.QSV M	0.932	0.863	0.952	0.858	0.067	0.047	0.136	0.952	0.861
	43	R.MNN	0.943	0.874	0.960	0.881	0.056	0.039	0.125	0.960	0.875
<b>GW O2</b>	44	R. CSVM	0.928	0.850	0.949	0.862	0.071	0.050	0.150	0.950	0.854
	45	R.WNN	0.924	0.835	0.945	0.856	0.075	0.054	0.164	0.947	0.842
	46	R.QSV M	0.927	0.846	0.948	0.859	0.072	0.051	0.154	0.949	0.850
	47	R.MNN	0.926	0.847	0.948	0.851	0.073	0.051	0.152	0.948	0.849
<b>Orig.</b>	48	R. CSVM	0.932	0.854	0.951	0.869	0.067	0.048	0.145	0.952	0.859
	49	R.WNN	0.932	0.852	0.951	0.869	0.067	0.048	0.147	0.952	0.858
	50	R.QSV M	0.932	0.856	0.951	0.872	0.067	0.048	0.143	0.952	0.862
	51	R.MNN	0.930 8	0.857 8	0.951 3	0.856 8	0.069 2	0.048 7	0.142 2	0.951 4	0.857 0

Due to the many models, the RAPS method was implemented along with the CRITIC method to identify the superior model for classifying skin cancer images. Initially, beneficial and nonbeneficial criteria should be addressed before RAPS implementation. The accuracy, sensitivity, specificity, precision, and F1 score are beneficial criteria, while the error rate, false positive rate, false negative rate, and negative predictive value are nonbeneficial criteria. Criteria weights were calculated by the CRITIC method using Equations 33-35, as shown in Table 5.

**Table 5** Criteria weights determined using the CRITIC method for selecting the optimum model for skin cancer classification.

	<b>ACC</b>	<b>SV</b>	<b>SP</b>	<b>PR</b>	<b>ER</b>	<b>FPR</b>	<b>FNR</b>	<b>NPV</b>	<b>FIS</b>
<b>Correlation coefficient</b>	1.321	1.427	1.349	1.203	2.672	2.719	2.901	1.579	1.512
<b>Weight</b>	0.079	0.086	0.081	0.072	0.160	0.163	0.174	0.095	0.091

To demonstrate the application of the RAPS method, Table 6 presents samples of method implementation through step 5, considering only the best and worst cases. Taking this into account, step 4 was executed according to Equation 14 to produce 0.18316 and 0.3023 for Qk (max) and Qh (min), respectively. Using Equation 20, the P value equals 0.8389. Steps 6 and 7 are presented in Table 7, where “A2” represents Model 2 and “A34” represents Model 34. The final rankings of the 51 models are listed in Table 8. The outcome results based on the proposed methodology are shown in Fig. 4.

**Table 6** An example of implementing the RAPS method for skin cancer classification for models 2 and 34.

	<b>Step</b>	<b>ACC</b>	<b>SV</b>	<b>SP</b>	<b>PR</b>	<b>ER</b>	<b>FPR</b>	<b>FNR</b>	<b>NPV</b>	<b>FIS</b>
Weight		0.079	0.086	0.081	0.072	0.160	0.163	0.174	0.095	0.091
Max/Min		0.945	0.899	0.962	0.908	0.038	0.038	0.100	0.883	0.903
A2		0.945	0.899	0.961	0.908	0.054	0.039	0.100	0.961	0.903

A34		0.869	0.748	0.907	0.760	0.130	0.092	0.251	0.908	0.754
A2	Step 1	0.999	1.000	0.999	1.000	0.704	0.987	1.000	0.919	1.000
A34		0.919	0.832	0.944	0.837	0.295	0.417	0.400	0.973	0.834
A2	Step 2	0.079	0.085	0.081	0.072	0.112	0.160	0.173	0.087	0.090
A34		0.072	0.071	0.076	0.060	0.047	0.067	0.069	0.092	0.075
Qmax	Step 3	0.079	0.086	0.081	0.072					0.091
Qmin						0.160	0.163	0.174	0.095	
A2Umax	Step 5	0.079	0.085	0.080	0.072					0.090
A2Umin						0.112	0.160	0.173	0.087	
A34Umax		0.072	0.071	0.076	0.060					0.834
A34Umin						0.047	0.067	0.069	0.092	

**Table 7** Ranking of the skin cancer classifications for models 2 and 34 based on the RAPS method.

Model	Qk(i)max	Qh(i)max	Pi	PSi	Rank
A2Umax	0.18313	0	0.79121	0.94313	1
A2Umin	0	0.27646	0		
A34Umax	0.07565	0	0.37880	0.45154	51
A34Umin	0	0.14213	0		

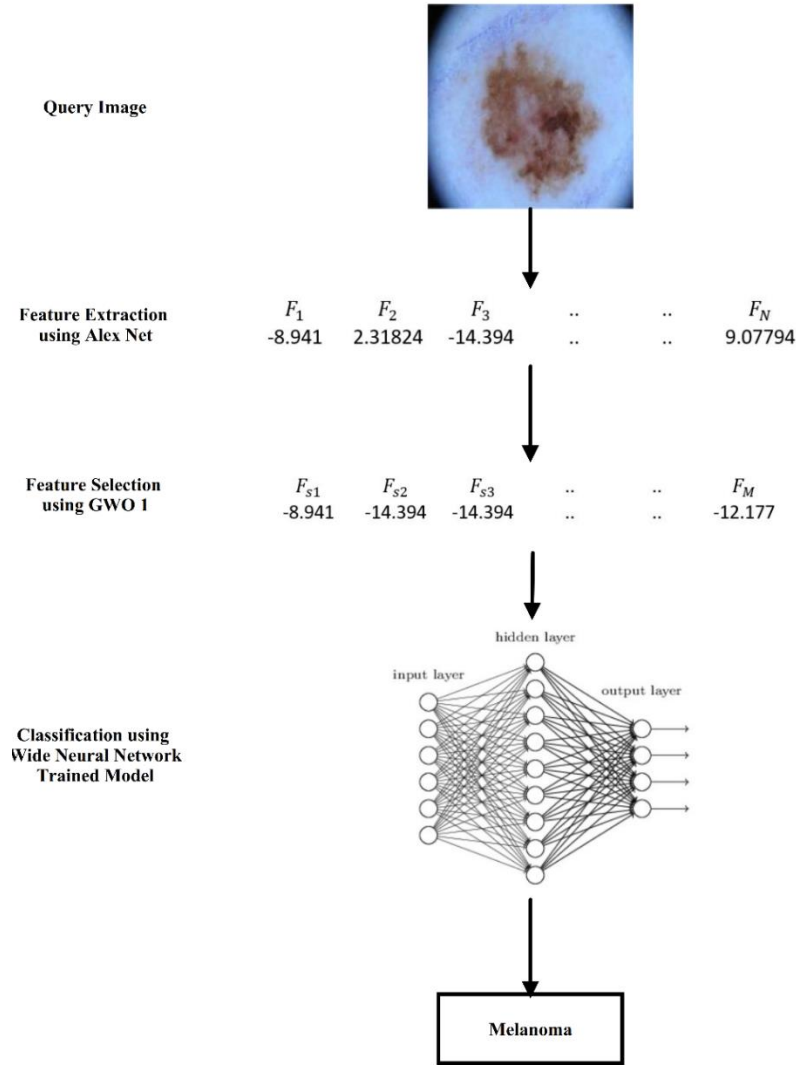
**Table 8** Final rankings of all the developed models for skin cancer classification based on the RAPS method.

Model no.	Model	PSi	Rank
A2	GWO1- AlexNet – WNN	0.943135	1
A10	Original-AlexNet – WNN	0.914284	2
A1	GWO1-AlexNet – CSVM	0.911944	3
A9	Original -AlexNet – CSVM	0.909893	4
A35	Original -Mobile V2 – CSVM	0.907412	5
A4	GWO 1- AlexNet – MNN	0.907079	6
A6	GWO 2- AlexNet – WNN	0.893249	7
A5	GWO 2- AlexNet – CSVM	0.893027	8
A43	GWO 1- ResNet 50 – QSVM	0.886682	9
A12	Original- AlexNet – MNN	0.885985	10
A3	GWO 1- AlexNet – QSVM	0.881655	11
A11	Original- AlexNet – QSVM	0.86874	12
A8	GWO 2- AlexNet-MNN	0.867611	13
A7	GWO 2- AlexNet – QSVM	0.854574	14
A42	GWO 1- ResNet 50 – MNN	0.81751	15
A50	Original- ResNet 50 – QSVM	0.809088	16
A51	Original - ResNet 50 – WNN	0.804455	17
A48	Original-ResNet 50 – CSVM	0.803863	18
A49	Original -ResNet 50 – LSVM	0.801991	19
A41	GWO 1- ResNet 50 - LSVM	0.79677	20
A23	Original- Inception V3 – QSVM	0.788236	21
A44	GWO 2- ResNet 50 - CSVM	0.78679	22
A40	GWO 1- ResNet 50 – CSVM	0.784624	23
A22	Original – Inception V3 – WNN	0.784136	24
A21	Original – Inception V3 – CSVM	0.782796	25
A47	GWO 2 – ResNet 50 – WNN	0.777952	26
A46	GWO 2 – ResNet 50 – QSVM	0.777777	27

A24	Original – Inception V3 – MNN	0.770182	28
A45	GWO 2 – ResNet 50 – LSVM	0.758805	29
A13	GWO 1 – Inception V3 – CSVM	0.756305	30
A16	GWO 1 – Inception V3 - MNN	0.751902	31
A14	GWO 1 – Inception V3 – WNN	0.751277	32
A15	GWO 1 – Inception V3 – QSVM	0.73575	33
A18	GWO 2 – Inception V3 – WNN	0.731325	34
A17	GWO 2 – Inception V3 – CSVM	0.715933	35
A19	GWO 2 – Inception V3 – QSVM	0.708347	36
A26	GWO 1 – Mobile V2 – WNN	0.645514	37
A28	GWO1 – MobileNet V2 – LSVM	0.640933	38
A36	Original – MobileNet V2 – WNN	0.640323	39
A25	GWO1 – MobileNet V2 – CSVM	0.63993	40
A38	Original – MobileNet V2 – LSVM	0.637589	41
A27	GWO 1 – MobileNet V2 – QSVM	0.637512	42
A37	Original – MobileNet V2 – QSVM	0.63563	43
A33	GWO 2- MobileNet V2 – LSVM	0.614532	44
A30	GWO 2- MobileNet V2 – CSVM	0.607877	45
A32	GWO 2 – MobileNet V2 – QSVM	0.607588	46
A31	GWO 2 – MobileNet V2 - WNN	0.603945	47
A20	GWO 2 – InceptionNet V3 – MNN	0.603502	48
A29	GWO 1 – MobileNet V2 - ESD	0.598146	49
A39	Original – MobileNet V2 - ESD	0.564224	50
A34	GWO 2 – MobileNet V2 – ESD	0.45154	51

As indicated in Table 6, the best model comprised GWO1 with AlexNet and a wide neural network, and the worst model was GWO2 with MobileNet V2 and an ensemble subspace discriminant. Although the results differ among the gray wolf optimizers, feature extractors, and classifiers, GWO 1, AlexNet, and the wide neural network achieved outstanding performances, with an accuracy of 94.5%, a sensitivity of 90%, and a specificity of 96%, as depicted in Table 4.

This can be interpreted as follows: compared to the complex architectures of its successors, the structure of AlexNet is relatively straightforward. This transparency allows easier interpretation of the extracted features, providing valuable insights into the model's decision-making process. The utilized dataset is limited in size compared to other image recognition tasks. AlexNet's ability to perform well with smaller datasets makes it a suitable choice for medical applications where data availability might be a constraint. In the challenge of accurate skin cancer classification, WNNs have yielded promising results, outperforming other models, such as LSVM, QSVM, CSVM, and ESD. While SVMs offer interpretability and robustness, they struggle with complex, nonlinear relationships often present in skin lesions. ESD tackles high dimensionality well but is limited by interpretability and cost. WNNs, however, shine with their ability to capture these intricate patterns, boasting greater accuracy thanks to their massive learning capacity and end-to-end learning. While challenges such as overfitting and interpretability linger, the sheer power and potential of WNNs make them the dominant force in this crucial medical application.



**Fig. 4** Overall summary of skin cancer classification using the proposed methodology.

When comparing our study's findings to those of related studies, numerous disparities emerge. Almost all the studies were differentiated according to the network utilized. In [18], the inception and residual networks were used; however, in [3], the authors modified Inception-ResNet V1 to detect skin cancer. Another network, namely, the stacked CNN, was used to detect melanoma classes [5]. A novel CNN based on VGG16 was developed to detect melanoma [29]. The same was conducted in [31] for developing a novel CNN for skin cancer classification. According to our proposed methodology, we used four CNN algorithms for feature selection and six ML algorithms for skin cancer classification. Therefore, the difference lies not only in the utilized algorithms but also in the modeling sequences. Additionally, we employed two versions of GWO to optimize feature selection, reduce the training time, and improve the classification process. The optimization context was adopted in [27] using the whale optimization algorithm, and the artificial rabbit optimizer was also used for another relevant work [4]. To the best of our knowledge, the MCDM approach has not been used to select the best model for skin cancer classification.

According to the proposed methodology, 51 models were found to represent skin cancer classification. A recent method called the RAPS was implemented to select the superior model. This study is the first to present a skin cancer classification model among 51 models, leading to the use of the MCDM to select a superior model.

## **Conclusions**

This study's conclusion involved the classification of skin cancer based on a variety of methods. The authors applied deep learning, machine learning, an optimization algorithm, and the MCDM to develop multiple models and select the best model. AlexNet, Inception V3, MobileNet V2, and ResNet 50 were employed as feature extractors. For feature selection, three scenarios were used: two with the GWO and one without an optimizer. Six different ML classifiers were used to classify skin cancer samples as basal cell carcinoma, melanoma, pigmented benign keratosis, or vascular lesions. Therefore, many models have been developed because of their distinct arrangements. A recent method called the RAPS was applied to select the optimum model for skin cancer classification. The arrangement of GWO1 with AlexNet and WNN yielded the best results. The proposed framework was tested on the ISIC 2017 dataset. The findings validate the applicability of the suggested framework. The development of multiple models for disease diagnosis provides a wide spectrum of models for precisely selecting the fittest model. Furthermore, feature reduction plays an influential role in the rapid detection and/or classification of skin cancer images. Even though feature reduction shortens training times, it does not necessarily guarantee the best performance of the developed models. Additionally, developing CAD systems can assist dermatologists in properly differentiating various classes of skin cancer, which in turn reduces errors and aids in providing correct treatment protocols. Future work can use other settings for feature extraction and selection, in addition to validation with other datasets for different or the same skin cancer classes.

**Competing interests:** The authors declare that they have no competing interests.

**Funding:** Not applicable for this study.

## **Data availability**

The datasets analyzed during the current study are available in the ISIC 2017 repository, <https://doi.org/10.1109/ISBI.2018.8363547>

## **CRedit authorship contribution statement**

Neven Saleh, Mohammed A Hassan, and Ahmed M. Salaheldin oversaw the study's conceptualization. Ahmed M. Salaheldin managed all aspects of the data acquisition, including procurement and preparation for the study. The introduction, background, and literature review were covered by Neven Saleh. Neven Saleh, Mohammed A Hassan, and Ahmed M. Salaheldin designed and developed the study's methodology. Neven Saleh analyzed the results to confirm the findings of the study. It was Neven Saleh and Ahmed M. Salaheldin who wrote and reviewed the article.

## Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this study, the authors partially used the QuillBot tool and a free trial of the American Journal Experts (AJE) (Springer Nature) platform for spelling and grammar checking. In addition, some sentences have been rephrased using these two services. After editing the text of the article, the authors reviewed and rewrote the contents of the article. The authors take full responsibility for the content of the article.

## References

1. Majumder S, Ullah MA. Feature extraction from dermoscopy images for melanoma diagnosis. *SN Appl. Sci.* 2019, 1(7): 753; <https://doi.org/10.1007/s42452-019-0786-8>
2. SM JPM, Aravindan C, Appavu R. Classification of skin cancer from dermoscopic images using deep neural network architectures. *Multimedia Tools and Applications.* 2023 ,82(10), 15763-15778; <https://doi.org/10.1007/s11042-022-13847-3>
3. Li Z, Chen Z, Che X, Wu Y, Huang D, Ma H, Dong Y. A classification method for multiclass skin damage images combining quantum computing and Inception-ResNet-V1. *Frontiers in Physics.* 2022 10, 1120; <https://doi.org/10.3389/fphy.2022.1046314>
4. Abd Elaziz M, Dahou A, Mabrouk A, El-Sappagh S, Aseeri, AO. An efficient artificial rabbits optimization based on mutation strategy for skin cancer prediction. *Computers in Biology and Medicine.* 2023, 163: 107154; <https://doi.org/10.1016/j.compbiomed.2023.107154>
5. Ahmed KT, Rustam F, Mehmood A, Ashraf I, Choi GS. Predicting skin cancer melanoma using stacked convolutional neural networks model. *Multimedia Tools and Applications.* 2023, 1-20; <https://doi.org/10.1007/s11042-023-15488-6>
6. Vestergaard ME, Macaskill PH M, Holt PE, Menzies SW. Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting. *British Journal of Dermatology.* 2008, 159(3):669-676.
7. Yu L, Chen H, Dou Q, Qin J, Heng PA. Automated melanoma recognition in dermoscopy images via very deep residual networks. *IEEE transactions on medical imaging.* 2017, 36(4), 994-1004; <https://doi.org/10.1109/TMI.2016.2642839>
8. Hatem MQ. Skin lesion classification system using a K-nearest neighbor algorithm. *Visual Computing for Industry, Biomedicine, and Art.* 2022, 5(1), 1-10; <https://doi.org/10.1186/s42492-022-00103-6>
9. Saleh N, Abdel Wahed M, Salaheldin AM. Transfer learning-based platform for detecting multi-classification retinal disorders using optical coherence tomography images. *International Journal of Imaging Systems and Technology.* 2022, 32(3):740-752; <https://doi.org/10.1002/ima.22673>
10. Bouzar-Benlabiod L, Harrar K, Yamoun L, Khodja MY, Akhloufi MA. A novel breast cancer detection architecture based on a CNN-CBR system for mammogram classification. *Computers in Biology and Medicine.* 2023, 163: 107133; <https://doi.org/10.1016/j.compbiomed.2023.107133>
11. Wang Z, Wu R, Xu Y, Liu Y, Chai R, Ma H. A two-stage CNN method for MRI image segmentation of prostate with lesion. *Biomedical Signal Processing and Control.* 2023, 82; <https://doi.org/10.1016/j.bspc.2023.104610>
12. Prakash NN, Rajesh V, Namakhwa, DL, Pande SD, Ahammad, SH. A DenseNet CNN-based liver lesion prediction and classification for future medical diagnosis. *Scientific African.* 2023, 20, e01629;

- [https://ui.adsabs.harvard.edu/link\\_gateway/2023SciAf..2001629P/doi:10.1016/j.sciaf.2023.e01629](https://ui.adsabs.harvard.edu/link_gateway/2023SciAf..2001629P/doi:10.1016/j.sciaf.2023.e01629)
13. Pandian R, Vedanarayanan V, Kumar DR, Rajakumar R. Detection and classification of lung cancer using CNN and Google net. *Measurement: Sensors*. 2022, 24, 100588; <http://dx.doi.org/10.1016/j.measen.2022.100588>
  14. Sudharshan PJ, Petitjean C, Spanhol F, Oliveira LE, Heutte L, Honeine P. Multiple instance learning for histopathological breast cancer image classification. *Expert Systems with Applications*. 2019, 117:103-111; <https://doi.org/10.1016/j.eswa.2018.09.049>
  15. Benkaddour MK, Bounoua A. Feature extraction and classification using deep convolutional neural networks, PCA and SVC for face recognition. *Traitement du signal*. 2017, 34: 77-91.
  16. Velliangiri S, Alagumuthukrishnan SJPCS. A review of dimensionality reduction techniques for efficient computation. *Procedia Computer Science*. 2019, 165: 104-111; <http://dx.doi.org/10.1016/j.procs.2020.01.079>
  17. Heidari AA, Mirjalili S, Faris H, Aljarah I, Mafarja M, Chen H. Harris hawks optimization: Algorithm and applications. *Future generation computer systems*. 2019, 97: 849-872; <https://doi.org/10.1016/j.future.2019.02.028>
  18. Singh SK, Banerjee S, Chakraborty A, Bandyopadhyay A. Classification of Melanoma Skin Cancer Using Inception-ResNet. In *Frontiers of ICT in Healthcare: Proceedings of EAIT 2022*, (pp. 65-74). Singapore: Springer Nature Singapore; [http://dx.doi.org/10.1007/978-981-19-5191-6\\_6](http://dx.doi.org/10.1007/978-981-19-5191-6_6)
  19. Nadimi-Shahraki MH, Zamani H, Asghari Varzaneh Z, Mirjalili S. A Systematic Review of the Whale Optimization Algorithm: Theoretical Foundation, Improvements, and Hybridizations. *Archives of Computational Methods in Engineering*. 2023, 1-47; <https://doi.org/10.1007/s11831-023-09928-7>
  20. Dorigo M, Stützle T *Ant colony optimization: overview and recent advances*. 2019, (pp. 311-351). Springer International Publishing.
  21. Pan WT. A new fruit fly optimization algorithm: taking the financial distress model as an example. *Knowledge-Based Systems*, 2012, 26, 69-74. <https://doi.org/10.1016/j.knsys.2011.07.001>
  22. Faris H, Aljarah I, Al-Betar MA, Mirjalili S. Gray wolf optimizer: a review of recent variants and applications. *Neural computing and applications*. 2018, 30: 413-435. <https://link.springer.com/article/10.1007/s00521-017-3272-5>
  23. Nurlitasari DA, Fuadah RYN, Magdalena R. Skin Cancer Classification Systems Using Convolutional Neural Network with Alexnet Architecture. In *Proceedings of the 2nd International Conference on Electronics, Biomedical Engineering, and Health Informatics: ICEBEHI 2021, 3-4 November, Surabaya, Indonesia* (pp. 227-236). Singapore: Springer Nature Singapore.
  24. Rashid J, Ishfaq M, Ali G, Saeed MR, Hussain M, Alkhalifah T, ..., Samand N. Skin cancer disease detection using transfer learning technique. *Applied Sciences*. 2022, 12(11), 5714; <https://doi.org/10.3390/app12115714>
  25. Abdulaal R, Bafail OA. Two New Approaches (RAMS-RATMI) in Multi-Criteria Decision-Making Tactics. *Journal of Mathematics*, 2022; <https://doi.org/10.1155/2022/6725318>
  26. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, (pp. 770-778).

27. Zhang L, Gao HJ, Zhang J, Badami B. Optimization of the convolutional neural networks for automatic detection of skin cancer. *Open Medicine*. 2020, 15(1): 27-37; <http://dx.doi.org/10.1515/med-2020-0006>
28. Dahou A, Aseeri AO, Mabrouk A, Ibrahim RA, Al-Betar MA, Elaziz MA. Optimal Skin Cancer Detection Model Using Transfer Learning and Dynamic-Opposite Hunger Games Search. *Diagnostics*. 2023, 13(9): 1579; <https://doi.org/10.3390/diagnostics13091579>
29. Tahir M, Naeem A, Malik H, Tanveer J, Naqvi RA, Lee SW. DSCC\_Net: Multi-Classification Deep Learning Models for Diagnosing of Skin Cancer Using Dermoscopic Images. *Cancers*. 2023, 15(7): 217-9; <https://doi.org/10.3390/cancers15072179>
30. Huynh AT, Hoang VD, Vu S, Le TT, Nguyen HD. Skin Cancer Classification Using Different Backbones of Convolutional Neural Networks. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. 2022, (pp. 160-172). Cham: Springer International Publishing.
31. Naeem A, Anees T, Fiza M, Naqvi RA, Lee SW. SCDNet: A Deep Learning-Based Framework for the Multiclassification of Skin Cancer Using Dermoscopy Images. *Sensors*, 2022, 22(15), 5652; <https://doi.org/10.3390/s22155652>
32. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O., ... , Farhan L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of big Data*. 2021, 8; <https://doi.org/10.1186/s40537-021-00444-8>
33. Li Z, Liu F, Yang W, Peng S, Zhou J. A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*. 2021; <https://doi.org/10.1109/TNNLS.2021.3084827>
34. Xu H. Comparison of CNN Models in Non-small Lung Cancer Diagnosis. In *2023 IEEE 3rd International Conference on Power, Electronics and Computer Applications (ICPECA)*, 2023, January, (pp. 1169-1174). IEEE.
35. Essel E, Lacy F, Elmedany W, Albaloochi F, Ismail Y. Driver Drowsiness Detection Using Fixed and Dynamic Thresholding. In *2022 International Conference on Data Analytics for Business and Industry (ICDABI)*, (2022, October), (pp. 552-557). IEEE.
36. Saleh N, Ali A, Ezzat O. Face Recognition-Based Smart Glass for Alzheimer's Patients. In *2022 10th International Japan-Africa Conference on Electronics, Communications, and Computations (JAC-ECC)*, (2022, December), (pp. 129-132). IEEE.
37. Palaz D, Magimai-Doss M, Collobert R. End-to-end acoustic modeling using convolutional neural networks for HMM-based automatic speech recognition. *Speech Communication*. 2019, 108: 15-32; <https://doi.org/10.1016/j.specom.2019.01.004>
38. Ayadi S, Lachiri, Z. (2022, March). Deep Neural Network for visual Emotion Recognition based on ResNet50 using Song-Speech characteristics. In *2022 5<sup>th</sup> International Conference on Advanced Systems and Emergent Technologies (IC\_ASET)* (pp. 363-368). IEEE.
39. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*. 2012, 25.
40. Junyue C, Zeebaree DQ, Qingfeng C, Zebari DA. Breast cancer diagnosis using hybrid AlexNet-ELM and chimp optimization algorithm evolved by Nelder-mead simplex approach. *Biomedical Signal Processing and Control*. 2023, 85, 105053; <http://dx.doi.org/10.1016/j.bspc.2023.105053>
41. Wang X, Li J, Tao J, Wu L, Mou C, Bai W, ... Deng Z. A Recognition Method of Ancient Architectures Based on the Improved Inception V3 Model. *Symmetry*. 2022, 14(12), 2679; <https://doi.org/10.3390/sym14122679>

42. Ali L, Alnajjar F, Jassmi HA, Gocho M, Khan W, Serhani MA. Performance evaluation of deep CNN-based crack detection and localization techniques for concrete structures. *Sensors*. 2021, 21(5), 1688; <https://doi.org/10.3390/s21051688>
43. Wamane N, Yadav A, Bhoir J, Shelke D, Kadam D. Comparison of Deep Learning Algorithms for Early Detection of Melanoma Skin Cancer on Dermoscopic and Nondermoscopic Images. In *International Conference on Multidisciplinary Trends in Artificial Intelligence*. (2023, June), (pp. 778-785). Cham: Springer Nature Switzerland.
44. Srinivasu PN, SivaSai JG, Ijaz MF, Bhoi AK, Kim W, Kang JJ. Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM. *Sensors*. 2021, 21(8), 2852; <https://doi.org/10.3390/s21082852>
45. Sae-Lim W, Wettayaprasit W, Aiyarak P. Convolutional neural networks using MobileNet for skin lesion classification. In *2019 16<sup>th</sup> international joint conference on computer science and software engineering (JCSSE)*, (2019, July), (pp. 242-247). IEEE.
46. Engelbrecht AP. *Fundamentals of computational swarm intelligence*. John Wiley & Sons, Inc; 2006.
47. Saleh N, Rosati S, Sharawi A, Wahed MA, Balestra G. An Optimal Scheduling for Medical Equipment Preventive Maintenance Over a Finite Planning Horizon Using Ant Colony Algorithm. *Journal of Clinical Engineering*. 2017, 42(3):142-149; <https://dx.doi.org/10.1097/JCE.0000000000000227>
48. Karaboga D. *An idea based on honeybee swarm for numerical optimization* (Vol. 200, pp. 1-10). Technical report-tr06, Erciyes university, engineering faculty, computer engineering department; 2005.
49. Hou Y, Gao H, Wang Z, Du C. Improved gray wolf optimization algorithm and application. *Sensors*. 2022, 22(10): 3810.
50. Mirjalili S, Mirjalili SM, Lewis A. Gray wolf optimizer. *Advances in engineering software*. 2014, 69: 46-61; <https://doi.org/10.1016/j.advengsoft.2013.12.007>
51. Zychlewicz M, Stanislawski R, Kaminski M. Gray wolf optimizer in design process of the recurrent wavelet neural controller applied for two-mass system. *Electronics*. 2022, 11(2): 177; <https://doi.org/10.3390/electronics11020177>
52. Panjaitan MI. Simple Additive Weighting (SAW) method in Determining Beneficiaries of Foundation Benefits. *Login: Jurnal Teknologi Komputer*. 2019, 13(1):19-25; <https://dx.doi.org/10.24224/login.v13i1.22>
53. Saleh N, Gaber MN, Eldosoky MA, Soliman AM. Vendor evaluation platform for acquisition of medical equipment based on multicriteria decision-making approach. *Scientific Reports*. 2023, 13:12746; <https://doi.org/10.1038/s41598-023-38902-3>
54. Taherdoost H, Madanchian M. Multicriteria decision making (MCDM) methods and concepts. *Encyclopedia*. 2023, 3(1): 77-87; <https://doi.org/10.3390/encyclopedia3010006>
55. Bafail OA, Abdulaal RM, Kabli MR. AHP-RAPS approach for evaluating the productivity of engineering departments at a public university. *Systems*. 2022, 10(4), 107; <https://doi.org/10.3390/systems10040107>
56. Urošević K, Gligorić Z, Miljanović I, Beljić Č, Gligorić M. Novel methods in multiple criteria decision-making process (Mcrat and raps)—Application in the mining industry. *Mathematics*. 2021, 9(16):1980; <https://doi.org/10.3390/math9161980>
57. Codella NC, Gutman D, Celebi ME, Helba B, Marchetti MA, Dusza SW, ... Halpern A. Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In *2018 IEEE 15th international symposium on biomedical imaging*

- (*ISBI* 2018), (2018, April), (pp. 168-172). IEEE;  
<https://doi.org/10.1109/ISBI.2018.8363547>
58. Saleh N, Yacoub KM, Salaheldin AM. Machine learning-based paradigm for the diagnosis of gestational diabetes In *2023 3rd International Conference on Electronic Engineering (ICEEM)*, (2023, October), (pp. 1-4). IEEE;  
<https://doi.org/10.1109/ICEEM58740.2023.10319496>
  59. Saleh N, Abdel Wahed M, Salaheldin AM. Computer-aided diagnosis system for retinal disorder classification using optical coherence tomography images. *Biomedical Engineering/Biomedizinische Technik*. 2022, 67(4): 283-294; <https://doi.org/10.1515/bmt-2021-0330>
  60. Keerthana D, Venugopal V, Nath MK, Mishra M. Hybrid convolutional neural networks with SVM classifiers for the classification of skin cancer. *Biomedical Engineering Advances*. 2023, 5, 100069; <https://doi.org/10.1016/j.bea.2022.100069>
  61. Tembhumne JV, Hebbar N, Patil HY, Diwan T. Skin cancer detection using an ensemble of machine learning and deep learning techniques. *Multimedia Tools and Applications*. 2023, 1-24; <http://dx.doi.org/10.1007/s11042-023-14697-3>
  62. Patil S, Jalan AK. Ensemble subspace discriminant classifiers for misalignment fault classification using vibro-acoustic sensor data fusion. *Journal of Vibration Engineering & Technologies*. 2022, 10(8):” 3169-3178; <https://doi.org/10.1007/s42417-022-00548-2>