Inteligencia Artificial 27(74), 102-116
doi: 10.4114/intartif.vol27iss74pp102-116

INTELIGENCIA ARTIFICIAL

http://journal.iberamia.org/

# Binary Classification of Skin Cancer Images Using Pre-trained Networks with I-GWO

Hadeer Hussein [1, *], Ahmed Magdy [1], Rehab F. Abdel-Kader [2], Khaled Abd El Salam [1, 3]

[1] Electrical Engineering Department, Suez Canal University, Ismailia, Egypt.
[2] Electrical Engineering Department, Port said University, Port said, Egypt.
[3] Department of Information System, College of Information Technology, Misr University for Science and Technology (MUST), 6th of October City 12566, Egypt.
* Corresponding author: Hadeer Hussein (e-mail: hadeerhussein@eng.suez.edu.eg).

**Abstract** One of the most prevalent forms of cancer worldwide is skin cancer. Determining disease characteristics necessitates a clinical evaluation of skin lesions, but this process is limited by long time horizons and a multiplicity of interpretations. Deep learning techniques have been created to help dermatologists with these issues as a higher patient survival rate depends on the early and precise detection of skin cancer. This research proposed a new approach for binary classification of dermoscopic images for skin cancer. The Improved Grey Wolf Optimizer (I-GWO) is used in this technique to fine-tune some hyperparameters' values of various pre-trained deep learning networks to maximize results. SqueezeNet, ShuffleNet, AlexNet, ResNet-18, and DarkNet-19 are the pre-trained networks that were employed. We tested the MED-NODE and DermIS databases in our investigation. Concerning the MED-NODE and DermIS datasets, the proposed method's highest accuracy results are 100% and 97%, respectively.

**Resumen** Una de las formas de cáncer más prevalentes en todo el mundo es el cáncer de piel. La determinación de las características de la enfermedad requiere una evaluación clínica de las lesiones cutáneas, pero este proceso está limitado por horizontes temporales prolongados y una multiplicidad de interpretaciones. Se han creado técnicas de aprendizaje profundo para ayudar a los dermatólogos con estos problemas, ya que una mayor tasa de supervivencia del paciente depende de la detección temprana y precisa del cáncer de piel. Esta investigación propuso un nuevo enfoque para la clasificación binaria de imágenes dermatoscópicas para el cáncer de piel. El Optimizador de lobo gris mejorado (I-GWO) se utiliza en esta técnica para ajustar los valores de algunos hiperparámetros de varias redes de aprendizaje profundo previamente entrenadas para maximizar los resultados. SqueezeNet, ShuffleNet, AlexNet, ResNet-18 y DarkNet-19 son las redes previamente entrenadas que se emplearon. Probamos las bases de datos MED-NODE y DermIS en nuestra investigación. Con respecto a los conjuntos de datos MED-NODE y DermIS, los resultados de mayor precisión del método propuesto son del 100% y el 97%, respectivamente.

**Keywords**: Binary classification, Deep learning, Dermoscopic images, I-GWO, Pre-trained networks, Skin cancer.
**Palabras clave**: Clasificación binaria, Aprendizaje profundo, Imágenes dermatoscópicas, I-GWO, Redes preentrenadas, Cáncer de piel.

## 1 Introduction

Computer-aided diagnostic (CAD) technologies are now required to assess and evaluate medical images in the modern-day [1]. Furthermore, CAD is critical for medical research, particularly in diagnostic and imaging radiography. When utilized correctly, the CAD system can lead to early illness identification, which can lead to early treatment choices, potentially saving lives [2]. For example, the capacity to identify a form of cancer at its early stages is closely connected to its effective detection and treatment [3]. Because cancer is a conglomeration of

**ISSN: 1137-3601 (print), 1988-3064 (on-line)**

several illnesses, early identification and therapy are critical [4]. Cancer is the largest cause of mortality among humans, according to several statistics. Skin cancer is a prevalent form; it often arises in the skin that has been exposed to sunlight regularly, however, cancer can occur elsewhere in the body. Skin cancer is highly noticeable because it begins in the epidermis, the uppermost skin layer. This shows that the CAD system can use photos of skin lesions without examining any other related information to make a preliminary diagnosis [5].

Melanoma is the most fatal kind of skin cancer discovered in humans, causing pigmented markings on moles to appear on the skin [6]. Melanoma is caused by any abnormalities in the melanin-producing cells, which give the skin its color. Certain risk factors for melanoma include a history of sunburn, a compromised immune system, pale skin, genetic factors, and unnecessary exposure to ultraviolet light [7]. Melanoma begins to develop and spread over the outer skin layer before infiltrating the inner layers, where it eventually links with the blood and lymph arteries. When skin cancer is recognized in its early stages, it has a higher chance of being treated than when it is discovered in its advanced stages. However, early detection of skin cancer is costly [8].

Determining whether a lesion is malignant or benign is challenging because skin lesions resemble one another. A normal mole is often the same color as the skin, such as brown, black, or tan, with a prominent border that distinguishes it from the nearby skin. Moles are typically round or oval and less than 0.25 inch in diameter. A search for moles with irregular borders, forms, colors, and moles larger than 0.25 inch in diameter is performed to find atypical mole features that might signal skin cancer or melanoma. Numerous strategies, such as genetic algorithms, Artificial Neural Networks (ANNs), and Convolutional Neural Networks (CNNs) have been proposed to analyze skin pain and categorize it as melanoma or benign [8]. All these approaches have been proven to be more cost-effective, efficient, and less painful than traditional medical procedures. However, in many computer vision applications, it is apparent that both CNNs and deep learning are the preferred techniques [9].

Pretrained Deep Learning Networks (PDLNs) are networks that have previously learned to extract powerful and useful characteristics from real images and utilize them as a starting point for learning a new task. They have been utilized to boost performance and lower computing costs in a variety of disciplines. In image classification tasks, PDLNs have been frequently employed [10].

The Grey Wolf Optimizer (GWO) is a meta-heuristic algorithm inspired by the natural hierarchy of grey wolves and their hunting behavior. It imitates the hunting processes of seeking prey, surrounding prey, and attacking prey. It gives competitive optimization performance outcomes. To address difficulties like local optima and population variety, several GWO versions have been proposed. Among these versions is I-GWO, a version with a novel search method that provides the global optimum. Several articles have presented the I-GWO method to improve optimization approaches [11], [12].

In this paper, we use five PDLNs: AlexNet, ResNet-18, SqueezeNet, ShuffleNet, and DarkNet-19 to detect and classify skin cancer diagnosis. The proposed method depends on using these PDLNs with I-GWO to enhance the accuracy of these PDLNs to 100%, using the MED-NODE dataset, by optimizing their parameters using I-GWO. The proposed technique will enable early identification and diagnosis of skin cancer. This, in turn, will allow for more efficient treatment and a reduction in the disease's death rate.

The rest of the paper is structured as follows. Section 2 presents related works of skin cancer classification and detection. Section 3 shows the used materials and methods. The experimental results are described in section 4. Finally, section 5 illustrates the conclusion of the research.

## 2 Related Works

This section lists various research that has been published in the field of skin cancer classification and detection. It focuses on recent works that have used Deep Learning (DL) for the same purpose.

Kwasigroch et al. [13] proposed using a CNN with hill climbing for search space to classify skin lesions. This method increased network size, lowering computing costs. Adegun et al. [14] described an encoder and decoder network with skip connections linking subnetworks. The proposed CNN was used to segment skin lesions and categorize them pixel by pixel. CNN, according to Song et al. [15], can segment, recognize, and categorize skin lesions. To control the imbalanced datasets, they employed a loss function based on the Jaccard distance and the focal loss. Manne et al. [16] proposed a CNN-based skin cancer classification system. They showed a completely automated computer technique for skin lesion classification. In this study, three models—ResNet-18, AlexNet, and VGG16—were pre-trained to act as feature generators. These recovered attributes are subsequently used to train support vector machines.

Thurnhofer-Hemsi and Domnguez [17] presented a CNN architecture for skin cancer diagnosis. They claimed that the DenseNet201 network's results are appropriate for this application. Kousis et al. [18] studied deep learning algorithms as well as a mobile app for accurate skin cancer screening. They proposed the XGBoost, an average of

the top 8 DL models, and an average of 15 DL models. Nawaz et al. [19] presented DL algorithms for melanoma diagnosis. CNN was utilized in this approach to extract visual characteristics. These attributes are then sent into two ANN models, the first of which is a CNN and was used to identify the target locations. In addition, the second NN is a recurrent CNN (RCNN), which detects the location of the lesion. Finally, the lesion segmentation was carried out using the Fuzzy K-means (FKM) method based on the established position. Reis et al. [20] employed CNNs to Skin Cancer Diagnosis (SCD) and identify lesion locations. In this model, the input images are pre-processed before being segmented using the UNet network. The lesion region is clipped depending on the segmentation results, and this segment is utilized as the input of a CNN model dubbed InSiNet to categorize the input image.

Tabrizchi et al. [21] used an upgraded CNN model based on the VGG-16 architecture of the visual geometry group to diagnose SC. In this study, the VGG model's design was changed such that it is more compatible with SCD disorders and can be identified with more accuracy than the initial model. Changes to filter dimensions and NN activation functions are included in this enhancement. Mazoure et al. [22] proposed DUNEScan, a web service for evaluating uncertainty in SCD using deep neural networks. Several CNN models, including ResNet50T, EfficientNet, Inceptionv3, and MobileNetv2, were utilized in this system to predict skin cancer, and diagnostic uncertainty was calculated using the following criteria: average, and variance of learning models. Mohakud and Dash [23] used a mix of CNN and GWO algorithms for SCD. In this procedure, the input images were initially pre-processed to eliminate duplicate information from the images, and the CNN training duration was also increased. The GWO technique was then used to tweak the convolutional neural network's hyperparameters. These hyperparameters provide the specifications of the convolution filters in the CNN model's layers. The best-discovered configuration was utilized for SCD in fresh samples.

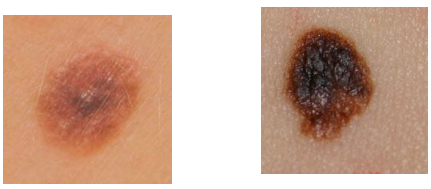
Shorfuzzaman [24] used a DL ensemble model to treat SCD. The transfer learning approach was utilized in this model, which contains numerous partial CNN classifications operating concurrently in the form of an ensemble system. Finally, the outcomes of these models were combined, and the final output was determined using an integration model. Magdy et al. [25] proposed two approaches for identifying and categorizing benign and malignant tumors in dermoscopic images. The first technique employs K-Nearest Neighbor (KNN) as a classifier, with PDLNs acting as feature extractors. The second technique optimizes its hyperparameters by combining AlexNet and GWO. The authors also used artificial neural networks, support vector machines, and convolutional neural networks to investigate two techniques to categorize skin cancer images: Machine Learning (ML) and DL. The experiments were carried out on 4000 images from the ISIC archive collection, and the proposed methods outperformed other evaluated approaches, with some models obtaining an accuracy of more than 99%.

# 3 Materials and Methods

This research proposed a method for skin cancer classification, evaluated various pre-trained models, and compared their results to the proposed one. This section displays the datasets utilized, the system model, and the proposed approach.

## 3.1 Datasets

In this work, we used two datasets: MED-NODE and DermIS datasets. MED-NODE dataset [26] (accessible at [27]) comprises images of skin lesions from the digital image archive of the Department of Dermatology of the University Medical Center Groningen (UMCG). It contains 170 colored JPEG images, divided into 70 melanoma and 100 nevus images. We scaled them to a specific dimension (227×227 pixels) due to their varying dimensions. They were then filtered using a median filter. This dataset was randomly divided into 90% for training and 10% for testing, resulting in 153 and 17 images, respectively. Figure 1 depicts samples from this dataset after scaling and filtering.



(a) (b)

Figure 1. Samples of MED-NODE dataset after scaling and filtering. (a) nevus; and (b) melanoma.

The DermIS dataset [28] (accessible at [29]) is a Dermatology Information System for dermoscopic images. This dataset was created in collaboration with the University of Erlangen's Department of Dermatology and the University of Heidelberg's Department of Clinical Social Medicine. It comprises 1000 colorful JPEG images with size of 600×450, 500 of which are benign and 500 of which are malignant. This dataset was randomly split into 90% training and 10% testing, yielding 900 and 100 images, respectively. Samples of this dataset are shown in Figure 2.

*(a)* *(b)*

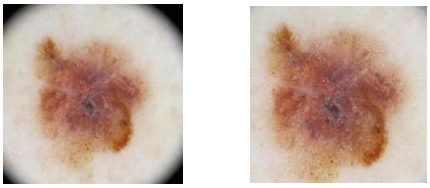
Figure 2. Samples of DermIS dataset. (a) benign; and (b) malignant.

## 3.2 Preprocessing

The data were pre-processed in the following manner. First, because images have varying dimensions, it is necessary to scale all images to a given size. All images have been scaled to 250×250 pixels. Second, images were filtered using the median filter. Third, the skin hair that occurred in the images was deleted using various morphological procedures so that it would not affect the classification results because it may be regarded as a part of the lesion, as shown in Figure 3. Figure 4 illustrates that images were trimmed to a proper dimension (150×150), as many photographs have black borders due to microscope use. Fifth, as shown in Figure 5 (b), we segmented the lesion in every image and removed the remaining images with black. Sixth, the black patches around the lesion were used to determine the region of interest, as shown in Figure 5 (c); nevertheless, this step altered the image sizes. Each image must be resized to the initial size (250×250). Figure 6 depicts all previously performed preprocessing procedures. In practice, the segmentation phase impacted classification accuracy and system performance, thus we avoided it to achieve better results.

(a) (b)

Figure 3. Sample of skin hair removal step. (a) before; and (b) after.



(a) (b)

Figure 4. Sample of Image trimming stage. (a) before; and (b) after.

(a) (b) (c)

Figure 5. Segmentation and crop of the lesion area. (a) Input image; (b) Segmented image; and (c) Segmented image with surround cropping.

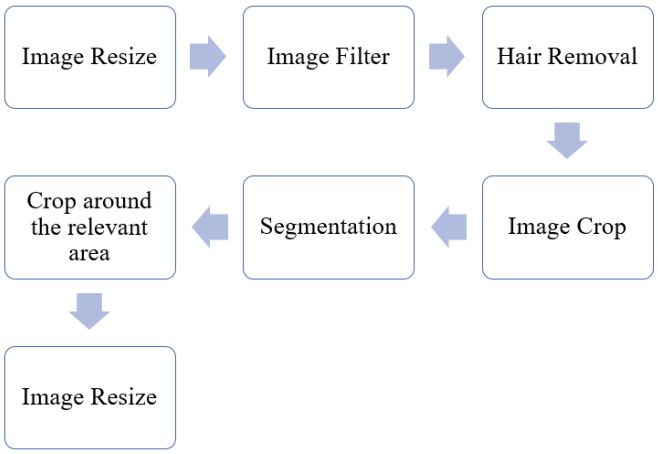



Figure 6. Steps of the dataset preprocessing.

## 3.3 System Model

In this paper, we tested five pre-trained networks as classifiers to classify images from two datasets. We also proposed a method that combines these pre-trained models with I-GWO and compared the results of PDLNs with the proposed method, as shown in Figure 7.

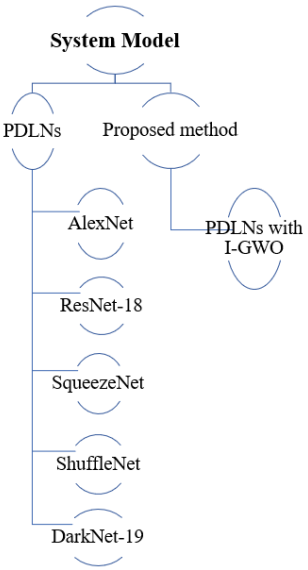



Figure 7. The system model contains PDLNs and the proposed method.

### 3.3.1 PDLNs

A pre-trained model is a stored network that has been previously trained on a large dataset, typically on a large-scale image-classification task. We employed five distinct PDLNs in this paper: AlexNet, ResNet-18, SqueezeNet, ShuffleNet, and DarkNet-19, which are mentioned below.

- **AlexNet**

AlexNet is a CNN model that largely affects deep learning applications in computer vision. It easily won the 2012 ImageNet LSVRC-2012 competition (15.3 percent botch rates versus 26.2 percent blunder rates in the runner-up, which is VGG-16). The organization's configuration was like Yann LeCun et alLeNet, 's but deeper, with more channels per layer and layered convolutional layers. Convolutions, maximum pooling, dropout, information growth, ReLU initiations, and stochastic gradient descent with force were all important. It adds ReLU initiations after each convolutional and completely related layer. Furthermore, instead of regularization, dropout is utilized to deal with overfitting [30].

- **ResNet-18**

The ILSRVC-2015 competition was won by the ResNet-50 model, which had a 3.57% error rate with an input image size of 224 × 224 pixels. Shaoqing Ren, Kaiming He, Jian Sun, and Xiangyu Zhang developed the well-known DL model ResNet. ResNet-18 contains 18 layers, whereas ResNet-50 has 50 layers, each having two or three convolutional layers. ResNet-101 is a 101-layer DL model [31].

- **SqueezeNet**

With several 2D convolutions, ReLU, max-pooling, and concatenation layers, SqueezeNet has 68 layers in total and 1.2 million learnable parameters. A dropout layer is also included to avoid overfitting. The pre-trained SqueezeNet network is retrained with over a million photos from the ImageNet collection of diverse objects. It features a wide feature set and a significantly simplified architecture that allows it to achieve higher accuracies with less computing cost and training time. SqueezeNet has demonstrated good performance with transfer learning in several investigations [32].

- **ShuffleNet**

Megvii Inc (also known as Face++) announced ShuffleNet, which they say is an incredibly computation-efficient CNN architecture optimized for mobile devices with 10-150 MFLOPs of computational capacity. To minimize computing costs while retaining accuracy, the ShuffleNet employs pointwise group convolution and channel shuffling. On ImageNet classification, it produces lower top-1 error than the MobileNet system and delivers a 13x real speedup over AlexNet while maintaining comparable accuracy [33].

- **DarkNet-19**

DarkNet-19 is a model of a convolutional neural network. To create predictions, this pre-trained model employs 1×1 filters to condense the feature representation between 3×3 convolutions and global average pooling. The DarkNet-19 is composed of 19 convolutional layers and five max-pooling layers that were trained using over one million photos from the ImageNet database. DarkNet-19 can classify images into over 1000 distinct things, including keyboards, mice, animals, and even humans [34].

Pretrained deep networks are tested as classifiers using Transfer Learning (TL) in this work. TL is a machine learning approach that repurposes a model developed for one task for another. It is commonly employed when there is a lack of training data. Data augmentation, on the other hand, can aid in overcoming the data difficulty. Because melanoma and benign lesions are so similar, it takes a long time to distinguish and classify them, we need transfer learning. Transfer learning is also more efficient in categorizing related lesions, making it the recommended strategy. Transfer learning networks are trained on massive datasets, and their model weights are frozen before the last few layers are changed for a different dataset [35].

A pre-trained network can be used to begin learning a new task. TL is much faster and simpler than training a network from scratch with randomly assigned weights. We started with data loading, preprocessing, and splitting. We then ran a pre-trained network. Convolutional layers of the network retrieved image features, which are the final learnable and final classification layers utilized to categorize the input picture. The classification layer sets the

output classes of the network. We removed the classification layer and replaced it with a new one that did not include class labels. To freeze the weights of the network's previous layers, we set their learning rates to zero. During training, the network did not alter the parameters of the frozen layers. Because the gradients of the frozen layers do not need to be computed, freezing the weights of multiple early layers can greatly expedite network training.

Each network requires different sizes of input images, so we employed an augmented image data store to automatically resize the training images. On the training images, we specified additional augmentation procedures: random reflection, translation, scaling, and rotation. The addition of data stops the network from overfitting and memorizing the exact features of the training pictures. After that, the training choices were given, as well as the number of epochs to be trained in. When applying transfer learning, it is not necessary to train for as many epochs. The fine-tuned network was utilized to classify the testing images and determine classification accuracy. Figure 8 depicts the procedures involved in using PDLNs as TL classifiers.

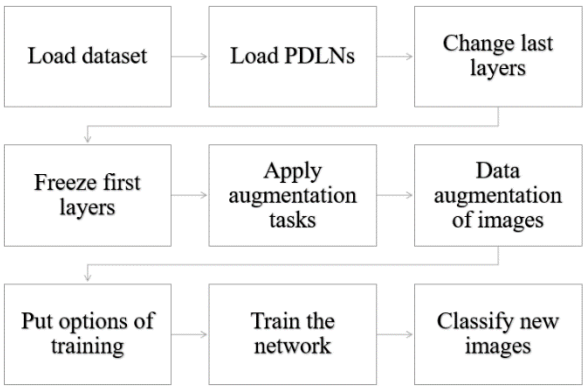



Figure 8. Process Steps of classification of pre-trained models with TL.

### 3.3.2 Proposed Method

In this paper, we proposed a method to detect and classify skin cancer disease, which is PDLNs with I-GWO. The GWO algorithm [11] is built around the concept of hunting. Wolves are members of a hunting pack that contains numerous grey wolves. Wolves in a pack are classed according to their leadership skills. There are four varieties of wolves in a pack: alpha ($\alpha$), beta ($\beta$), delta ($\delta$), and omega ($\omega$). The group's choice is made by the leader of the hunting procedure. The surviving wolves' dominance progressively declines in the following order: $\beta$, $\delta$, and $\omega$. Such wolves hunt and report back to their master's about their better situations. In this phase, which is the course of the pursuit, the first grey wolves search for and contain prey $\alpha$, $\beta$, and $\delta$ have a better understanding of potential prey locations for mathematical models of hunting behavior [36].

The algorithm offers both huge benefits and possible drawbacks. One of the benefits of the GWO method is its ease of implementation. The algorithm's fundamentals are quite straightforward, making it simple to learn and execute. Second, the GWO method can do extensive global searches. The system, which is based on the prey-seeking behavior of grey wolf packs, can successfully explore probable optimum solutions over the whole search space.

This enables it to identify globally optimum solutions to unconstrained continuous optimization problems. Furthermore, the GWO method has a high rate of convergence. By imitating wolves' collaborative and competitive behavior to update the solution vector, the method can swiftly converge to a better solution, thereby speeding up the optimization process. Finally, the GWO algorithm has some parameter adaptability. It may automatically alter the parameter values during the search process to increase the resilience and performance of the algorithm. This flexibility may adapt to the peculiarities of various challenges, reducing the difficulty of parameter modification. However, the GWO algorithm has certain possible downsides. First, it is more sensitive to the problem's restrictions. When dealing with limited optimization issues, additional processing is necessary to verify that the outcomes are feasible. Second, while the GWO algorithm can do global searches, it may nevertheless fall back on local optimum solutions in some complicated issues. Finally, parameter tweaking in the algorithm is a challenge; various issues may necessitate different parameter choices, and experimentation and debugging are required to get better performance [37].

In GWO, $\alpha$, $\beta$, and $\delta$ direct $\omega$ wolves to parts of the search space where they are likely to locate the best answer. This behavior may be trapped in a locally optimum solution. Another negative consequence is a decline in population variety, which causes GWO to fall into the local optimum. To address these limitations, I-GWO was suggested as a GWO upgrade. The I-GWO enhances wolf-hunting search strategies by employing a novel search approach known as Dimension Learning-based Hunting (DLH). The DLH search approach is inspired by wolves' hunting behavior in the wild, and it broadens the scope of global search through multi-neighbor learning. The I-GWO then has both candidate wolves created by the DLH and the GWO seeks techniques in each iteration to relocate the wolf $X_i$ from the present location to a better place. Furthermore, the I-GWO employs an extra choosing and updating phase in each iteration to pick the winning candidate wolf and update the current position for the following iteration. The I-GWO comprises three phases: initializing, moving, and selecting and updating, as shown below [12].

*Initializing phase:* N wolves are randomly dispersed in the search space within a certain range $[l_i, u_j]$ by (1).

$$X_{ij} = l_j + rand_j[0,1] \times (u_j - l_j), i \in [1, N], j \in [1, D] \quad (1)$$

The i-th wolf's location in the t-th iteration is represented as a vector of real numbers $X_i(t) = \{x_{i1}, x_{i2}, ..., x_{iD}\}$, where D is the problem's dimension number. The whole wolf population is kept in a Pop matrix with N rows and D columns. The fitness function, $f(X_i(t))$, determines the fitness value of $X_i(t)$.

*Movement phase:* individual hunting, in addition to group hunting, is an intriguing social activity of grey wolves, and it is our drive to enhance GWO. DLH is an extra movement strategy approach that is incorporated by I-GWO. Each wolf in DLH is recognized by its neighbors as another contender for the new post of $X_i(t)$.

*Selecting and updating phase:* During this phase, the best candidate is chosen by comparing the fitness values of two candidates, $X_{i\text{-}GWO}(t+1)$ and $X_{i\text{-}DLH}(t+1)$, using (2).

$$X_i(t+1) = \begin{cases} X_{i-GWO}(t+1), & if\ f(X_{i-GWO}) < f(X_{i-DLH}) \\ X_{i-DLH}(t+1), & otherwise \end{cases} \quad (2)$$

$X_{i\text{-}GWO}$ is the first candidate for the new position of wolf $X_i(t)$, and it is calculated by (3).

$$X_{i-GWO}(t+1) = \frac{X_{i1}(t) + X_{i2}(t) + X_{i3}(t)}{3} \quad (3)$$

In addition to $X_{i\text{-}GWO}(t+1)$, $X_{i\text{-}DLH}(t+1)$ is another candidate for the new position of wolf $X_i(t)$ that is generated by the DLH search strategy which is computed by (4).

$$X_{i-DLH,d}(t+1) = X_{i,d}(t) + rand \times (X_{n,d}(t) - X_{r,d}(t)) \quad (4)$$

, where the d-th dimension of $X_{i\text{-}DLH,d}(t+1)$ is computed by taking the d-th dimension of a random neighbor $X_{n,d}(t)$ chosen from $N_i(t)$, the neighbors of $X_i(t)$ shown at (5), and a random wolf $X_{r,d}(t)$ chosen from Pop.

$$N_i(t) = \{X_j(t) | D_i\left(X_i(t), X_j(t)\right) \leq R_i(t), X_j(t) \in Pop\} \quad (5)$$

, where $D_i$ represents the Euclidean distance between $X_i(t)$ and $X_j(t)$. $R_i(t)$ is a radius determined by (6) utilizing the Euclidean distance between $X_i(t)$ 's present position and the candidate point $X_{i\text{-}GWO}(t+1)$.

$$R_i(t) = \|X_i(t) - X_{i-GWO}(t+1)\| \quad (6)$$

After that, to update the new position of $X_i(t+1)$, if the fitness value of the selected candidate is smaller than $X_i(t)$, the picked candidate updates $X_i(t)$. Otherwise, $X_i(t)$ in Pop remains unaffected. After completing this operation for all persons, the iteration counter is incremented by one, and the search can be repeated until the predetermined number of iterations is achieved [12].

In this paper, I-GWO is chosen to be used as an optimizer with PDLNs. The same stages as for pre-trained models as classifiers were followed, which is shown in Figure 8, except that variables at the mini-batch size and initial learning rate hyperparameters were added to the training choices to be optimized by I-GWO, as illustrated in Figure 5. We also selected the maximum number of I-GWO iterations and the number of search agents. The lower

and upper limits for the required parameters are then determined. In detail, we initialized alpha, beta, and delta positions, initialized search agent positions, returned search agents that go beyond the search space boundaries, calculated the objective function for each search agent, updated alpha, beta, and delta positions, and updated the position of search agents including omegas. Figure 9 depicts the procedure steps of the proposed method.

# 4 Experimental Results

This section compares PDLN results and the proposed method, PDLNs with I-GWO, results. These experiments are tested on MED-NODE and DermIS datasets. System implementation and used performance metrics are also shown in this section.

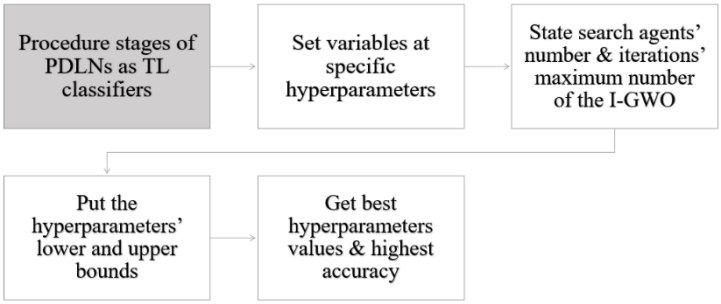



Figure 9. Process Steps of the proposed method.

## 4.1 System Implementation

The frameworks that have been implemented have been tested and reviewed on the following software and hardware configurations:

- Operating system: Windows 10 Pro.
- Compiler: MATLAB R2020b.
- Processor: Intel (R) Core (TM) i7-9750H CPU @ 2.60GHz 2.59 GHz.
- Installed RAM: 16.0 GB (15.9 GB usable).
- System type: 64-bit operating system, x64-based processor.

## 4.2 Performance Metrics

To estimate the performance of the PDLNs, we employed five measurements (see (7), (8), (9), (10), and (11) below).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{7}$$

$$\text{F1 score} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}} \tag{8}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{9}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{10}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{11}$$

- *TP (True Positive):* The number of malignant or melanoma samples correctly classified.

- *TN (True Negative):* The number of benign or nevus samples correctly classified.
- *FP (False Positive*): The number of benign or nevus samples that were mistakenly diagnosed as malignant or melanoma.
- *FN (False Negative):* The number of malignant or melanoma samples that were mistakenly diagnosed as benign or nevus.

## 4.3 Results and Discussion

This section describes the findings and parameters utilized in the tested networks as well as the proposed technique.

AlexNet, ResNet-18, SqueezeNet, ShuffleNet, and DarkNet-19 were the five PDLNs we examined. These models were evaluated as TL classifiers. Because each network requires a distinct size of the input image, we used an augmented image data store to resize the training images automatically. We specified other augmentation techniques on the training images, which are randomly reflection along the horizontal axis, randomly translation up to 30 pixels, horizontally scaling and vertically up to 10%, and rotation up to 30 pixels. The training hyperparameters of PDLNs are as follows: 50 maximum epochs, 32 mini-batch sizes, and $1\times10^{-4}$ initial learning rate.

When employing PDLNs with I-GWO (proposed technique), the number of search agents in I-GWO was limited to 5, with a maximum of 20 iterations. I-GWO was used to adjust two PDLNs' hyperparameters: mini-batch size and initial learning rate. The lower and upper limitations for mini-batch size were set to 2 and 64, respectively. The initial learning rate has lower and upper limitations of $1\times10^{-5}$ and $1\times10^{-3}$, respectively.

### 4.3.1 In MED-NODE Dataset

In the first Dataset, MED-NODE, we split it into 90% for training (153 images) and 10% for testing (17 images). The best hyperparameters of I-GWO that give the best accuracy are shown in Table 1. These hyperparameters provide the best accuracy for PDLNs with I-GWO (100%) compared to them as TL classifiers, as shown in Table 2. In Table 2, When using PDLNs as TL classifiers, AlexNet and ResNet-18 have the highest accuracy (94.1176%) compared to other networks, but all networks achieve 100% in the proposed method. The confusion matrix of AlexNet and the training progress of ResNet-18 are shown in Figure 10 and Figure 11, respectively.

Table 1. The best hyperparameters that I-GWO optimized for PDLNs (in MED-NODE dataset).

| PDLNs | Mini Batch Size | Initial Learning Rate |
|---|---|---|
| AlexNet | 28.57590344 | 2.38E-04 |
| ResNet-18 | 37.94404237 | 6.08E-04 |
| SqueezeNet | 10.19685983 | 3.77E-04 |
| ShuffleNet | 17.25031776 | 5.47E-04 |
| DarkNet-19 | 5.159421059 | 1.97E-04 |

Table 2. Performance of PDLNs as TL classifiers vs. performance of the proposed method (using MED-NODE dataset).

| PDLNs | PDLNs as TL classifiers | | | | | The proposed method | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Precision (%) | F1 score (%) | Sensitivity (%) | Specificity (%) | Accuracy (%) | Precision (%) | F1 score (%) | Sensitivity (%) | Specificity (%) | Accuracy (%) |
| AlexNet | 100 | 94.7368 | 90 | 100 | 94.1176 | 100 | 100 | 100 | 100 | 100 |
| ResNet-18 | 100 | 94.7368 | 90 | 100 | 94.1176 | 100 | 100 | 100 | 100 | 100 |
| SqueezeNet | 90 | 90 | 90 | 85.7143 | 88.2353 | 100 | 100 | 100 | 100 | 100 |
| ShuffleNet | 90 | 90 | 90 | 85.7143 | 88.2353 | 100 | 100 | 100 | 100 | 100 |
| DarkNet-19 | 100 | 88.8889 | 80 | 100 | 88.2353 | 100 | 100 | 100 | 100 | 100 |

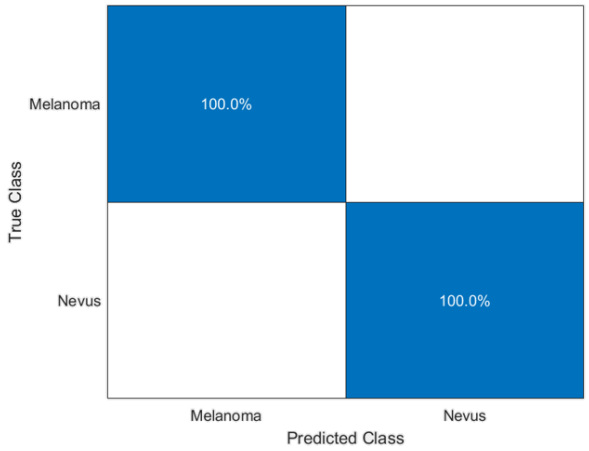



Figure 10. Confusion matrix of AlexNet in proposed method (using MED-NODE dataset).


Figure 11. Training Progress of ResNet-18 in proposed method (using MED-NODE dataset).

### 4.3.2 In DermIS Dataset

In the second Dataset, DermIS, we split into 90% for training (900 images) and 10% for testing (100 images). The best hyperparameters of I-GWO that give the best accuracy are shown in Table 3. These hyperparameters provide the best accuracy for PDLNs with I-GWO (97%) compared to them as TL classifiers, as shown in Table 4. In Table 4, When using PDLNs as TL classifiers, AlexNet has the highest accuracy (94%) compared to other networks, but it achieves 96% and SqueezeNet achieves 97% in the proposed method. The confusion matrix of ShuffleNet and the training progress of DarkNet-19 are shown in Figure 12 and Figure 13, respectively.

Table 3. The best hyperparameters that I-GWO optimized for PDLNs (in the DermIS dataset).

| PDLNs | Mini Batch Size | Initial Learning Rate |
|---|---|---|
| AlexNet | 32.94150309 | 4.66E-04 |
| ResNet-18 | 55.24170352 | 9.98E-04 |
| SqueezeNet | 23.90725418 | 8.76E-04 |
| ShuffleNet | 29.05948951 | 9.36E-04 |
| DarkNet-19 | 37.90979144 | 9.66E-04 |

Table 4. Performance of PDLNs as TL classifiers vs. performance of the proposed method (using DermIS dataset).

| PDLNs | PDLNs as TL classifiers | | | | | The proposed method | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Precision (%) | F1 score (%) | Sensitivity (%) | Specificity (%) | Accuracy (%) | Precision (%) | F1 score (%) | Sensitivity (%) | Specificity (%) | Accuracy (%) |
| AlexNet | 97.83 | 93.75 | 90 | 98 | 94 | 100 | 95.0495 | 96 | 100 | 96 |
| ResNet-18 | 93.62 | 90.72 | 88 | 94 | 91 | 97.0588 | 91.6667 | 95 | 98 | 94 |
| SqueezeNet | 97.78 | 92.63 | 88 | 98 | 93 | 100 | 93.2039 | 96 | 100 | 97 |
| ShuffleNet | 95.74 | 92.78 | 90 | 96 | 93 | 97.7778 | 93.75 | 90 | 98 | 94 |
| DarkNet-19 | 97.73 | 91.49 | 86 | 98 | 92 | 97.8723 | 94.8454 | 92 | 98 | 95 |

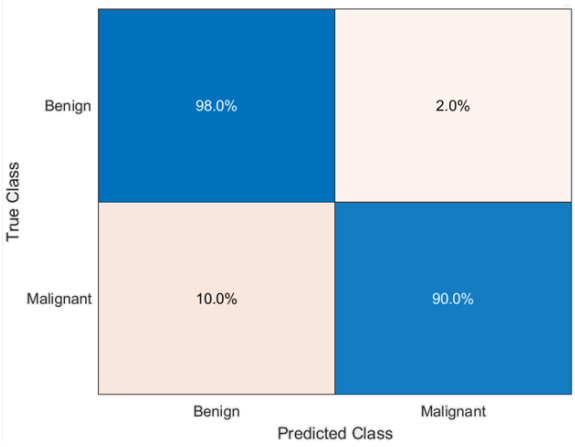



Figure 12. Confusion matrix of ShuffleNet in the proposed method (using DermIS dataset).

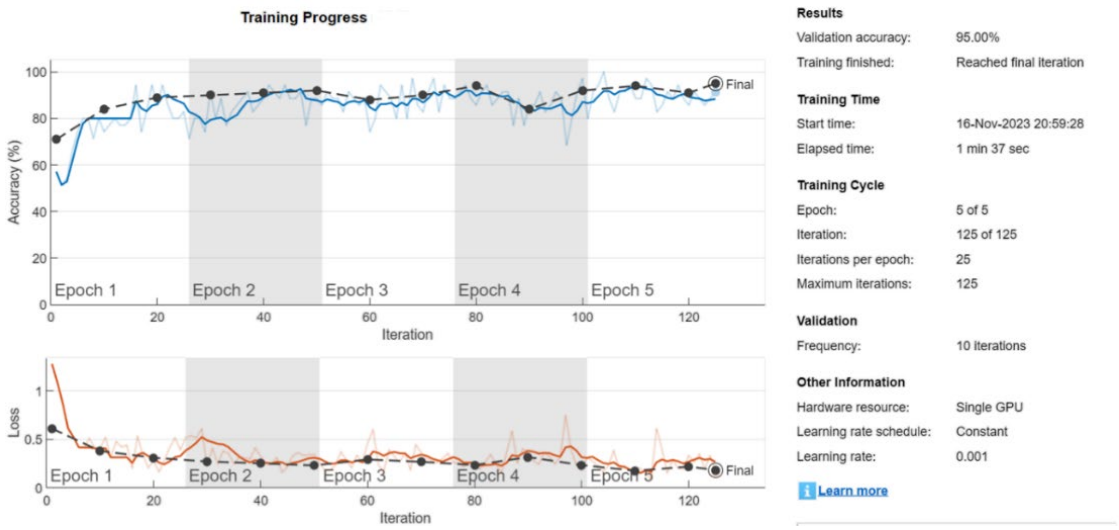



Figure 13. Training Progress of DarkNet-19 in the proposed method (using DermIS dataset).

# 5 Conclusion

One of the worst malignancies in the world is skin cancer. Early detection and diagnosis of skin lesions is critical for determining the appropriate therapy for the patient and, in the event of malignant lesions, increasing the patient's chance of life. This condition is diagnosed manually by competent dermatologists, but it takes time and is challenging. This technique may be made quicker, faster, and more precise by utilizing CAD technologies. A new approach for detecting and classifying skin cancer was proposed in this paper. This approach was proposed to detect and categorize benign and malignant tumors in dermoscopic images. The proposed method employs a combination of an optimization methodology (I-GWO) and PDLNs to optimize PDLNs' hyperparameters. It was tested on samples from two databases, MED-NODE and DermIS. According to the results, the proposed technique can detect skin cancer in the MED-NODE and DermIS datasets with 100% and 97% accuracy, respectively.

The limitation in our work is the long processing time of I-GWO and some PDLNs which prevents us from using large datasets. Future work will involve training and testing the models on much bigger datasets with multiple classes. Metadata, such as age, location of lesion, thickness of lesion, and progression of lesion, may be included with datasets to enhance classifier metrics. Furthermore, the use of real-time skin cancer detection models with mobile applications can considerably increase accessibility and convenience for both patients and healthcare providers.

# Acknowledgments

First of all, I am extremely grateful to Allah for giving me the capability and the power to achieve this dissertation. Then, I would like to express my greatest appreciation, gratitude, and thanks to my supervisors: Prof. Rehab Farouk Abd El Kader, Dr. Khaled Abd El Salam Ali, and Dr. Ahmed Magdy Mohamed for their ceaseless encouragement and guidance. Finally, I am really thankful to my family, friends, and colleagues for their love, understanding, support, and patience who were powerfully inspiring me during this period of my life. Without them, I could never have reached this step.

# References

[1] A. Masood and A. Ali Al-Jumaily, 'Computer Aided Diagnostic Support System for Skin Cancer: A Review of Techniques and Algorithms', *Int J Biomed Imaging*, vol. 2013, pp. 1–22, 2013, doi: 10.1155/2013/323268.

[2] A. M. Abdel-Zaher and A. M. Eldeib, 'Breast cancer classification using deep belief networks', *Expert Syst Appl*, vol. 46, pp. 139–144, Mar. 2016, doi: 10.1016/j.eswa.2015.10.015.

[3] J. D. Wulfkuhle, L. A. Liotta, and E. F. Petricoin, 'Proteomic applications for the early detection of cancer', *Nat Rev Cancer*, vol. 3, no. 4, pp. 267–275, Apr. 2003, doi: 10.1038/nrc1043.
[4] Y.-E. Choi, J.-W. Kwak, and J. W. Park, 'Nanotechnology for Early Cancer Detection', *Sensors*, vol. 10, no. 1, pp. 428–455, Jan. 2010, doi: 10.3390/s100100428.
[5] H. E. Kanavy and M. R. Gerstenblith, 'Ultraviolet Radiation and Melanoma', *Semin Cutan Med Surg*, vol. 30, no. 4, pp. 222–228, Dec. 2011, doi: 10.1016/j.sder.2011.08.003.
[6] B. G. Goldstein and A. O. Goldstein, 'Diagnosis and management of malignant melanoma', *Am Fam Physician*, vol. 63, no. 7, pp. 1359–1369, 2001.
[7] Y. Goncharova, E. A. S. Attia, K. Souid, and I. V Vasilenko, 'Dermoscopic Features of Facial Pigmented Skin Lesions', *ISRN Dermatol*, vol. 2013, pp. 1–7, Feb. 2013, doi: 10.1155/2013/546813.
[8] A. Esteva *et al.*, 'Dermatologist-level classification of skin cancer with deep neural networks', *Nature*, vol. 542, no. 7639, pp. 115–118, Jan. 2017, doi: 10.1038/nature21056.
[9] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, 'Deep learning for visual understanding: A review', *Neurocomputing*, vol. 187, pp. 27–48, Apr. 2016, doi: 10.1016/j.neucom.2015.09.116.
[10] X. Han *et al.*, 'Pre-trained models: Past, present and future', *AI Open*, vol. 2, pp. 225–250, 2021, doi: 10.1016/j.aiopen.2021.08.002.
[11] S. Mirjalili, S. M. Mirjalili, and A. Lewis, 'Grey Wolf Optimizer', *Advances in Engineering Software*, vol. 69, pp. 46–61, Mar. 2014, doi: 10.1016/j.advengsoft.2013.12.007.
[12] M. H. Nadimi-Shahraki, S. Taghian, and S. Mirjalili, 'An improved grey wolf optimizer for solving engineering problems', *Expert Syst Appl*, vol. 166, p. 113917, Mar. 2021, doi: 10.1016/j.eswa.2020.113917.
[13] A. Kwasigroch, M. Grochowski, and A. Mikolajczyk, 'Neural Architecture Search for Skin Lesion Classification', *IEEE Access*, vol. 8, pp. 9061–9071, 2020, doi: 10.1109/access.2020.2964424.
[14] A. A. Adegun and S. Viriri, 'Deep Learning-Based System for Automatic Melanoma Detection', *IEEE Access*, vol. 8, pp. 7160–7172, 2020, doi: 10.1109/access.2019.2962812.
[15] L. Song, J. Lin, Z. J. Wang, and H. Wang, 'An End-to-End Multi-Task Deep Learning Framework for Skin Lesion Analysis', *IEEE J Biomed Health Inform*, vol. 24, no. 10, pp. 2912–2921, Oct. 2020, doi: 10.1109/jbhi.2020.2973614.
[16] R. Manne, S. Kantheti, and S. Kantheti, 'Classification of Skin cancer using deep learning, convolutional neural Networks-Opportunities, and vulnerabilities-A Systematic Review', *International Journal for Modern Trends in Science and Technology, ISSN*, vol. 6, pp. 2455–3778, 2020.
[17] K. Thurnhofer-Hemsi and E. Domínguez, 'A Convolutional Neural Network Framework for Accurate Skin Cancer Detection', *Neural Process Lett*, vol. 53, no. 5, pp. 3073–3093, Oct. 2020, doi: 10.1007/s11063-020-10364-y.
[18] I. Kousis, I. Perikos, I. Hatzilygeroudis, and M. Virvou, 'Deep Learning Methods for Accurate Skin Cancer Recognition and Mobile Application', *Electronics (Basel)*, vol. 11, no. 9, p. 1294, Apr. 2022, doi: 10.3390/electronics11091294.
[19] M. Nawaz *et al.*, 'Skin cancer detection from dermoscopic images using deep learning and fuzzy k-means clustering', *Microsc Res Tech*, vol. 85, no. 1, pp. 339–351, Aug. 2021, doi: 10.1002/jemt.23908.
[20] H. C. Reis, V. Turk, K. Khoshelham, and S. Kaya, 'InSiNet: a deep convolutional approach to skin cancer detection and segmentation', *Medical & Biological Engineering & Computing*, vol. 60, no. 3, pp. 643–662, Jan. 2022, doi: 10.1007/s11517-021-02473-0.
[21] H. Tabrizchi, S. Parvizpour, and J. Razmara, 'An Improved VGG Model for Skin Cancer Detection', *Neural Process Lett*, vol. 55, no. 4, pp. 3715–3732, Jul. 2022, doi: 10.1007/s11063-022-10927-1.
[22] B. Mazoure, A. Mazoure, J. Bédard, and V. Makarenkov, 'DUNEScan: a web server for uncertainty estimation in skin cancer detection with deep neural networks', *Sci Rep*, vol. 12, no. 1, Jan. 2022, doi: 10.1038/s41598-021-03889-2.
[23] R. Mohakud and R. Dash, 'Designing a grey wolf optimization based hyper-parameter optimized convolutional neural network classifier for skin cancer detection', *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 8, pp. 6280–6291, Sep. 2022, doi 10.1016/j.jksuci.2021.05.012.
[24] M. Shorfuzzaman, 'An explainable stacked ensemble of deep learning models for improved melanoma skin cancer detection', *Multimedia Syst*, vol. 28, no. 4, pp. 1309–1323, Apr. 2021, doi: 10.1007/s00530-021-00787-5.
[25] A. Magdy, H. Hussein, R. F. Abdel-Kader, and K. A. El Salam, 'Performance Enhancement of Skin Cancer Classification Using Computer Vision', *IEEE Access*, vol. 11, pp. 72120–72133, 2023, doi: 10.1109/access.2023.3294974.

[26] I. Giotis, N. Molders, S. Land, M. Biehl, M. F. Jonkman, and N. Petkov, 'MED-NODE: A computer-assisted melanoma diagnosis system using non-dermoscopic images', *Expert Syst Appl*, vol. 42, no. 19, pp. 6578–6585, Nov. 2015, doi: 10.1016/j.eswa.2015.04.034.

[27] MED-NODE, 'Dermatology database used in MED-NODE'. [Online]. Available: https://www.cs.rug.nl/imaging/databases/melanoma_naevi/index.html

[28] DermIS, 'Dermatology Information System'. [Online]. Available: https://www.dermis.net/dermisroot/en/home/index.htm

[29] DermIS, 'Skin Lesion Dermis Dataset'. [Online]. Available: https://www.kaggle.com/datasets/farhatullah8398/skin-lesion-dermis-dataset

[30] S. Suganyadevi and S. V, 'DarkCVNet: Optimized Pneumonia and Covid 19 detection using CXR Images', Jul. 2022, doi: 10.21203/rs.3.rs-1294582/v1.

[31] K. Vyshnavi, A. Kintali, K. Sravan, V. Kothuru, and V. Gowtham, 'Performance Analysis of Pneumonia Detection Using ResNets Technique'. Jan. 2021. doi: 10.51397/OAIJSE09.2021.0002.

[32] A. Saad, A. Usman, S. Arif, and M. Liwicki, 'Bearing Fault Detection Scheme Using Machine Learning for Condition Monitoring Applications', in *ICMAME 2023 Conference Proceedings*, in ICMAME 2023. ECER, 2023. doi: 10.53375/icmame.2023.137.

[33] X. Zhang, X. Zhou, M. Lin, and J. Sun, 'ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices', in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2018.

[34] B. T. Hung, V. B. Semwal, N. Gaud, and V. Bijalwan, 'Violent Video Detection by Pre-trained Model and CNN-LSTM Approach', in *Proceedings of Integrated Intelligence Enable Networks and Computing*, Springer Singapore, 2021, pp. 979–989. doi: 10.1007/978-981-33-6307-6_99.

[35] S. Jain, U. Singhania, B. Tripathy, E. A. Nasr, M. K. Aboudaif, and A. K. Kamrani, 'Deep Learning-Based Transfer Learning for Classification of Skin Cancer', *Sensors*, vol. 21, no. 23, p. 8142, Dec. 2021, doi: 10.3390/s21238142.

[36] A. Guernine and M. T. Kimour, 'Optimized Training for Convolutional Neural Network Using Enhanced Grey Wolf Optimization Algorithm', *Informatica*, vol. 45, no. 5, Aug. 2021, doi: 10.31449/inf.v45i5.3497.

[37] Z. Pang, Y. Wang, and F. Yang, 'Application of Optimized Kalman Filtering in Target Tracking Based on Improved Gray Wolf Algorithm', Oct. 2023, doi: 10.21203/rs.3.rs-3455700/v1.