Optimizing Hyperparameters for Improved Melanoma Classification using Metaheuristic Algorithm

Shamsuddeen Adamu[1], Hitham Alhussian[2], Dr. Norshakirah Aziz[3], Dr. Said Jadid Abdulkadir[4], Dr. Ayed Alwadin[5], Dr. Abdullahi Abubakar Imam[6], Aliyu Garba[7], Yahaya Saidu[8]

Computer Information Science Department, Universiti Teknologi PETRONAS, Seri Iskandar, Malaysia[1, 2, 3, 4, 7]
IAIICT, Ahmadu Bello University, Zaria, Nigeria[1, 7]
Universiti Brunei Darussalam (UBD) Brunei, Darussalam[6]
Computer Science Department Community College, King Saud University, Riyadh, Saudi Arabia[5]

***Abstract*—Melanoma, a prevalent and formidable skin cancer, necessitates early detection for improved survival rates. The rising incidence of melanoma poses significant challenges to healthcare systems worldwide. While deep neural networks offer the potential for precise melanoma classification, the optimization of hyperparameters remains a major obstacle. This paper introduces a groundbreaking approach that harnesses the Manta Rays Foraging Optimizer (MRFO) to empower melanoma classification. MRFO efficiently fine-tunes hyperparameters for a Convolutional Neural Network (CNN) using the ISIC 2019 dataset, which comprises 776 images (438 melanoma, 338 non-melanoma). The proposed cost-effective DenseNet121 model surpasses other optimization methods in various metrics during training, testing, and validation. It achieves an impressive accuracy of 99.26%, an AUC of 99.56%, an F1 score of 0.9091, a precision of 94.06%, and a recall of 87.96%. Comparative analysis with EfficientB1, EfficientB7, EfficientNetV2B0, NesNetLarge, ResNet50, VGG16, and VGG19 models demonstrates its superiority. These findings underscore the potential of the novel MRFO-based approach in achieving superior accuracy for melanoma classification. The proposed method has the potential to be a valuable tool for early detection and improved patient outcomes.**

Keywords*—Deep learning; machine learning; classification; metaheuristic algorithm; CNN

I. Introduction

Skin cancer is characterized by the uncontrolled growth of abnormal cells in the skin's outermost layer (epidermis) due to DNA damage and mutations. Melanoma, a highly aggressive form of skin cancer, is on the rise, posing significant challenges for healthcare systems in terms of early detection and intervention [1]. The precise cause of all melanomas remains unclear, but exposure to ultraviolet (UV) radiation from sunlight or tanning lamps and beds significantly heightens the risk of developing this form of skin cancer [2]. In the United States, the American Cancer Society's estimations predict that approximately 97,610 new cases of melanoma will be diagnosed in the year 2023 [3]. This contrasts with the year 2020, during which a total of 324,635 individuals tested positive for melanoma, with an unfortunate outcome in 57,043 cases, resulting in fatalities [4]. The five-year relative survival rate for melanoma is 93%, with a range of 99% for cases diagnosed at a localized stage (83% of cases) to 27% for cases diagnosed at a faraway stage (4%) [5]. Nevertheless, due to the inherent challenges associated with early detection, both by professionals and patients, there is a pressing need to develop an effective method for the early detection of melanoma and skin cancer. To ensure precise and timely diagnosis, accurate equipment is essential, even in the hands of highly skilled professionals. Dermatologists typically initiate the diagnostic process by employing non-invasive dermoscopy techniques on suspicious areas. However, this conventional method primarily relies on visual examination by the naked eye for preliminary diagnosis. This reliance on visual assessment can introduce challenges, as it can be cumbersome, susceptible to inaccuracies, and dependent on subjective interpretation in a clinical environment [6]. Given these inherent limitations, there is an urgent need to integrate advanced diagnostic tools and technologies to improve the accuracy and efficiency of early melanoma detection.

Deep learning (DL), with its multi-layered artificial neural networks, has ushered in a transformative era in image classification [7]. At the forefront of this revolution are Convolutional Neural Networks (CNNs), renowned for their unparalleled ability to autonomously discern intricate features from raw pixel data [8]. Through iterative training on extensive datasets, DL models refine their parameters to excel in identifying objects and patterns, finding application in diverse domains such as medical imaging [9], autonomous driving [10]. Moreover, in the realm of natural language processing [11]. The exceptional precision and effectiveness of CNNs, particularly in analyzing complex medical images, continue to inspire researchers to leverage them in addressing intricate challenges across various fields.

CNNs have emerged as powerful tools, and researchers often leverage predefined CNN models such as VGG16, VGG19, ResNet, DenseNet, Inception, etc., which come pre-trained on extensive image datasets. These pre-trained models serve as valuable starting points for various image-related tasks, as they have already learned a rich set of features from massive amounts of data. Achieving optimal performance with these models requires careful consideration of hyperparameters, and proper selection of these hyperparameters has a direct impact on the performance of the model [12]. Effective hyperparameter selection stands as a critical endeavor for optimizing a CNN's performance in various applications [13], particularly in the realm of medical imaging, where precise and accurate results are paramount

[14]. Unlocking the full potential of CNNs in this critical domain hinges on the careful process of hyperparameter optimization. To address this challenge, metaheuristic algorithms have emerged as invaluable tools for efficiently searching for optimal hyperparameter values [15]. In contrast to traditional gradient-based optimization methods, such as stochastic gradient descent (SGD), which may struggle with the complex and non-convex landscapes [16] often encountered in medical image analysis, metaheuristics excel. These algorithms, which encompass a range of techniques, including genetic algorithms, particle swarm optimization, simulated annealing, and ant colony optimization, among others, have demonstrated their adaptability and efficiency in exploring and exploiting intricate parameter spaces [17]. In medical imaging, where accuracy and speed are paramount, metaheuristic algorithms offer a promising approach to enhance the capabilities of DL models, ultimately contributing to more accurate disease diagnosis and treatment planning [18].

This study centers on hyperparameter optimization for melanoma classification, employing the DenseNet-121 transfer learning model in tandem with the MRFO metaheuristic algorithm. The principal objective is to elevate melanoma detection accuracy, a pivotal factor in enabling early diagnosis and ultimately improving patient outcomes.

In this study, we make the following significant contributions:

- Novel CNN-Based Framework: A novel CNN framework, developed for reliable classification of Melanoma using the ISIC2019 dataset.
- Transfer Learning Utilization: We utilize ten pre-trained models for transfer learning, benefiting from their knowledge learned on extensive datasets to enhance melanoma classification performance.
- Hyperparameter Optimization: To boost classification performance, we fine-tune hyperparameters for both the CNN model and pre-trained models using the Manta Rays Foraging Optimizer (MRFO) algorithm. This process identifies optimal configurations for each pre-trained model, enhancing melanoma detection performance.

The subsequent sections of this paper are structured as follows:

Section II provides a comprehensive literature review on melanoma classification, offering insights into existing research and methodologies. Section III, a detailed overview of our proposed methodology, outlining the approach to melanoma classification using the CNN framework and the MRFO algorithm. Section IV delves into the experimental results and discussions, shedding light on findings and their implications. Finally, Section V presents conclusions and outlines potential avenues for future research in the field of melanoma classification.

II. Literature Review

In recent years, there has been a significant surge in the development of algorithms designed for the automated detection of melanoma using dermoscopy images. In the early 2000s, most automated melanoma classification solutions primarily relied on the utilization of manually crafted, low-level features such as shape, color, and texture [19]. However, recent studies, exemplified by [20], have signaled a shift in melanoma identification and recognition methodologies. This transition signifies a departure from the heavy reliance on manual feature engineering, marking a substantial evolution in the field. Presently, there is a noticeable uptick in the adoption of DL approaches for automated skin image analysis [21]. This paradigm shift reflects a move away from manual feature engineering towards methods that harness the capabilities of neural networks to autonomously learn and extract pertinent features from dermoscopy images. This transformation underscores the changing landscape in melanoma classification.

Daghrir et al. [22] introduced an innovative melanoma identification approach by combining three techniques using a majority voting method. Their method incorporates CNN along with two traditional ML approaches, SVM and KNN. These models were trained to recognize specific skin cancer features, including edges, texture, and color characteristics. While this ensemble approach improved overall performance, the combination of results from three different techniques can introduce computational complexity and potentially result in slower processing times. Nevertheless, the results demonstrated that CNN achieved the highest accuracy at 85.5%, followed by SVM with 71.8% accuracy and KNN with 57.3% accuracy.

To address the critical challenge of accurately classifying early Melanoma detection, Golnoori et al. in [23] introduced a novel approach by tackling the persistent complexity of selecting optimal neural network architectures and hyperparameters by employing metaheuristic optimization algorithms to fine-tune both pre-trained and scratch-trained CNN models. These optimized models' deep features were effectively combined and utilized to train a K-nearest neighbors (KNN) classifier. The results of their method demonstrated exceptional performance, achieving an accuracy of 81.6% with an F1-score of 80.9% on the ISIC 2017 dataset and 90.1% accuracy with an 89.8% F1-score on the ISIC 2018 dataset.

Segmentation plays a pivotal role in ML. In their study [24], the authors introduced a technique involving image segmentation. This technique utilizes anisotropic diffusion filtering and a rapid bounding box approach, followed by feature extraction through a hybrid feature extractor (HFE) and a feature extractor based on CNNs. The fusion of these extracted features facilitated the development of a highly accurate classification model capable of distinguishing between melanoma and non-melanoma images. The evaluation, conducted on two datasets, yielded outstanding results, including an accuracy rate of 99.85%, sensitivity of 91.65%, and specificity of 95.70%. These outcomes

underscore the remarkable effectiveness of this approach, surpassing the performance of previous ML algorithms.

To tackle the issue of automated segmentation of melanoma regions, the authors in [25] proposed a DL method based on a deep region-based Convolutional Neural Network (RCNN). This method comprises three main steps: skin refinement, melanoma region localization, and segmentation. The research was evaluated using a dataset consisting of 900 training images and 376 testing images from ISIC 2016 melanoma images. The results highlight the superiority of this approach compared to state-of-the-art techniques, achieving exceptional performance across various evaluation metrics: accuracy (0.94), specificity (0.94), sensitivity (0.97), F1 score (0.96), dice score (0.94), and Jaccard coefficient (0.93).

Ian their study [21], the authors comprehensively analyze the use of pre-trained CNN architectures for melanoma classification. They address a gap in previous research by investigating the specific features extracted by different CNN models. To enhance feature extraction, they introduce boundary localization to preserve critical skin lesion regions. They assess the effectiveness of eight pre-trained CNN models for deep feature extraction from these regions and employ various classifiers for melanoma detection. Across datasets like Ph2, ISIC 2016, ISIC 2017, and HAM10000, their approach achieves high accuracies of 98.33%, 80.47%, 81.16%, and 81%, outperforming state-of-the-art CNN methods.

The authors in [26] introduced a comprehensive three-phase framework for melanoma diagnosis, incorporating data segmentation via an extended Chan-Vese method, data augmentation, and CNN training facilitated by an active learning mini-batch process. To gauge its performance, they conducted an evaluation comparing it with established models such as DenseNet, InceptionV3, MobileNet, NASNet, and Xception. The assessment leveraged Standard Deviation (STD) as a metric to gauge model robustness and stability. The proposed active learning query strategy outperformed baseline methods in both performance and convergence across sixteen image datasets, underscoring the framework's effectiveness in accurate melanoma diagnosis. Nevertheless, the reliance on the active learning mini-batch process may not be universally applicable and may exhibit reduced effectiveness when dealing with imbalanced or noisy datasets.

Adepu et al. [27] present a novel approach for melanoma classification using a lightweight Deep-CNN-based framework. Their methodology incorporates knowledge distillation, Cost-Sensitive Learning with Focal Loss, and in-painting algorithms to enhance classification performance. To improve the model's performance, they introduce new CutOut variations and utilize test time augmentation as regularizers. On the ISIC-2020 dataset, the authors achieve a state-of-the-art result with a sensitivity of 81%, an AUC of 93, and a specificity of 90%. However, as the model processes a larger number of augmented images during each epoch, the use of CutOut may significantly increase the computational cost of training.

The authors in [28] introduce a CAD system utilizing the Online Region-based Active Contour Model (ORACM) to extract the Region of Interest (ROI) from skin lesions. The system demonstrates remarkable performance with an accuracy of 92.24% and perfect specificity and sensitivity (100%). However, it exhibits limitations, such as a heavy reliance on handcrafted features, sensitivity to the initial contour location and manual parameter selection, and the potential for the Non-dominated Sorting Genetic Algorithm (NSGA II) to find local instead of global optimum solutions, resulting in increased computational complexity.

In the realm of lightweight transfer learning models suitable for mobile devices, the study by the authors in [29] introduced the utilization of MobileNetV2 for the classification of melanoma images into benign and malignant categories. Their experiments conducted on multiple melanoma datasets yielded promising results, demonstrating the effectiveness of the proposed method with accuracy rates reaching as high as 85%, outperforming other network architectures. Additionally, the suggested architectural design of the head model, featuring a global average pooling layer followed by two fully-connected layers, not only contributed to high accuracy but also preserved the network's efficiency

The diagnosis of cancer during surgical treatment primarily relies on cancer stage or tumor thickness. In their study[30], the authors introduced two distinct methods aimed at classifying melanoma into two stages: the first stage includes stage 1 and stage 2, while the second encompasses stage 1, stage 2, or stage 3 melanoma. The proposed system employs a CNN, utilizing the Similarity Measure for Text Processing (SMTP) as the loss function. The experimental results include a comparison of various loss functions against the proposed SMTP loss function, demonstrating the superior efficiency of the proposed algorithm when contrasted with several other loss functions tailored for classification tasks.

The authors in [31] introduces a hybrid learning approach for melanoma detection, employing image processing techniques to enhance detection by addressing common issues such as hair, air bubbles, and image noise in dermoscopic images. The study presents an innovative hybrid model, merging DL and machine learning (ML), in its debut for melanoma detection. Performance evaluation on the Hamm 1000 (ISIC 2018) and ISIC 2020 datasets achieves exceptional accuracy scores, reaching 99.44% and 100%, respectively. Notably, the effectiveness of the saturation masking and wavelet transform techniques used may be sensitive to image quality, especially in cases of low resolution, artifacts, or variations in lighting, potentially impacting their noise reduction and lesion prominence enhancement capabilities.

In the study conducted by Nancy [32] presents a two-tier hybrid dual CNN (2-HDCNN) feature fusion approach for malignant melanoma prediction. It first identifies challenging samples, generates a Baseline Segregated Dataset (BSD), and preprocesses it. The second-tier CNN produces bottleneck features, combined with ABCD rule-derived features. These hybrid features are used with various classifiers, resulting in high accuracy (92.15%), precision (96.96%), specificity (96.8%), sensitivity (86.48%), and an AUC of 0.96 for diagnosing malignant melanoma on the ISIC 2018 dataset.

In conclusion, our review of the existing literature in melanoma classification reveals a significant gap in research focused on the optimization of hyperparameters and network structures. While numerous studies have explored ML and DL models for melanoma detection, very few have delved into the critical area of hyperparameter optimization using metaheuristic algorithms. Notably, none of the studies have investigated the potential of the MRFO algorithm in this context. This identified research gap highlights the unexplored potential of MRFO and other metaheuristic algorithms for enhancing the accuracy and efficiency of melanoma classification models.

III. Methodology

A. Introduction

This study introduces a comprehensive framework for melanoma classification Exploiting CNNs alongside transfer learning techniques and MRFO for hyperparameter optimization. The workflow consisting of data acquisition, pre-processing, splitting, classification, and optimization. These phases collectively enable accurate and reliable melanoma classification.

B. Dataset Description

The ISIC-2019 dataset [33], comprising 25,331 dermoscopy images, is used to evaluate the proposed technique. These images, categorized into eight distinct classes representing different types of skin lesions, exhibit varying dimensions, ranging from 600x450 to 1024x1024 pixels. In our evaluation, a total of 776 images are employed, including 438 melanoma and 338 non-melanoma cases.

C. DenseNet-121 Architecture

DenseNet121, a CNN architecture renowned for its effectiveness in visual object recognition, achieves state-of-the-art results with reduced parameters by intricately connecting each layer to both preceding and subsequent layers [34].

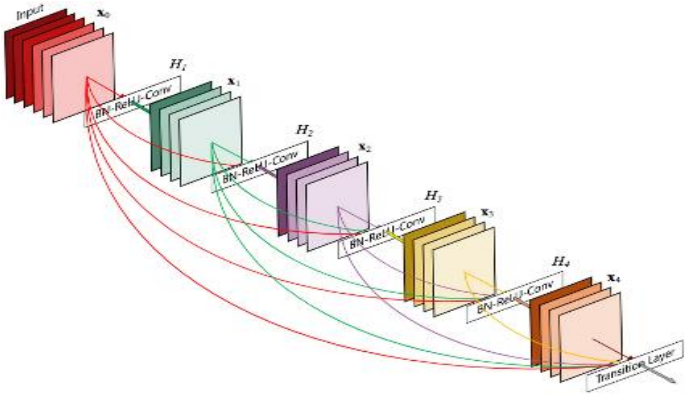

Fig. 1. DenseNet121 architecture.

DenseNet-121, one of several variants within the DenseNet family, is characterized by five convolution and pooling layers, three transition layers (at depths 6, 12, and 24), one classification layer (at depth 16), and two dense blocks employing 1x1 and 3x3 convolutions. Fig. 1 shows DenseNet121 architecture.

Total Layers = 5 + (6 + 12 + 24 + 16) × 2 = 121

In total, DenseNet-121 comprises 121 layers, making it a versatile choice for various computer vision tasks. This study specifically employs DenseNet-121 due to its compelling advantages, such as mitigating the vanishing-gradient problem, enhancing feature propagation, promoting feature reuse, and notably reducing the overall model parameter count [35]

D. Hyperparameter Selection

The study focuses on optimizing several critical hyperparameters to enhance model performance:

- Loss Function: Defines the loss metric used for model optimization.
- Training Batch Size: Determines the number of samples used in each training iteration.
- Model Dropout Ratio: Regulates the extent of regularization during training.
- Transfer Learning Ratio: Controls the extent of pre-trained weight utilization.
- Optimizer: Selects the optimization algorithm.
- Rotation Angle: Diversifies training data by simulating different lesion orientations.
- Shifts (Width and Height): Handles variations in lesion position and image composition.
- Zoom Level: Allows the model to learn features at various scales within images.
- Shear Transformation: Introduces controlled deformations to enhance the model's adaptability.
- Flipping: Simulates mirrored images (Horizontal and Vertical).

These hyperparameters collectively contribute to the adaptability, generalization, and optimization of the model for melanoma classification. Next steps involve leveraging the MRFO [36]. Metaheuristic Algorithm to fine-tune these hyperparameters systematically, ensuring that the model achieves the best possible results in the challenging task of classifying melanoma.

E. Manta Ray Foraging Metaheuristic Algorithm

The MRFO [36] algorithm incorporates three foraging behaviors: chain foraging, cyclone foraging, and somersault foraging. These behaviors are mathematically modeled as follows:

- Chain foraging: - In the MRFO [36] Chain foraging: Manta rays use their ability to detect plankton concentration to navigate towards better positions. The algorithm assumes the highest plankton concentration represents the best solution found so far. Manta rays form a foraging chain, moving towards both the food and the individual in front of them.

$$x_i^d(t+1) = \begin{cases} x_i^d(t) + r * (x_{i-1}^d(t) - x_i^d(t)) + \\ \alpha * \left(x_{best}^d(t) - x_i^d(t)\right) i = 1 \\ \\ x_i^d(t) + r * (x_{i-1}^d(t) - x_i^d(t)) + \\ \alpha * \left(x_{best}^d(t) - x_i^d(t)\right) i = 2, \dots N \end{cases} \quad (1)$$

$$\alpha = 2 * r * \sqrt{|\log(r)|} \quad (2)$$

In dimension d[th], $x_i^d(t)$ represents the i[th] individual's position at time (t) with r as a random vector between 0 and 1, α as a weighting coefficient, and $x_{best}^d(t)$ denoting a high-plankton concentration area. Fig. 2 illustrates the two-dimensional foraging behavior of individuals. An individual's position update depends on the previous (i-1)[th] individual's current position, $x_{i-1}^d(t)$, and the food source position, $x_{best}^d(t)$.

- Cyclone foraging: - Manta rays display a distinctive foraging behavior in which they form a long chain and spiral toward plankton patches in deep waters. This spiral foraging strategy resembles that seen in WOA [37]. In the case of cyclone foraging, manta ray swarms spiral towards food while also moving toward the ray in front, forming a line that takes on a spiral shape. Fig. 3 illustrates this cyclone foraging behavior in a 2-dimensional context. Swarm members follow the leader, traversing a spiral path toward food. The mathematical equation for modeling this two-dimensional spiral movement of manta rays is as follows:

$$\begin{cases} X_i(t+1) = X_{best} + r * \left(X_{i-1}(t) - X_i(t)\right) + \\ \quad e^{bw} * \cos(2\pi\omega) * \left(X_{best} - X_i(t)\right) \\ \\ X_i(t+1) = X_{best} + r * \left(X_{i-1}(t) - X_i(t)\right) + \\ \quad e^{bw} * \cos(2\pi\omega) * \left(X_{best} - X_i(t)\right) \end{cases} \quad (3)$$

Where, w is a random number between 0 and 1.

This behavior can extend to n-dimensional space. The mathematical model for cyclone foraging is defined as follows:

$$x_i^d(t+1) = \begin{cases} x_{best}^d + r * \left(x_{best}^d(t) - x_i^d(t)\right) + \\ \beta * \left(x_{best}^d(t) - x_i^d(t)\right) i = 1 \\ \\ x_{best}^d + r * \left(x_{i-1}^d(t) - x_i^d(t)\right) + \\ \beta * \left(x_{best}^d(t) - x_i^d(t)\right) i = 2, \dots n \end{cases} \quad (4)$$

$$\beta = 2e^{r1\frac{T-t+1}{T}} * \sin(2\pi r_1) \quad (5)$$

In the cyclone foraging strategy, individuals perform random searches relative to the food source as their reference point, promoting both exploitation and exploration. To encourage exploration, each individual is directed to seek a new position in the search space, away from the current best position, using a new random reference point. This approach enhances exploration, enabling the MRFO [36] algorithm to perform a broad global search. The mathematical equation is as follows:

β is the weight coefficient, T is the maximum number of iterations, and r_1 is a random number in [0, 1].

$$x_{rand}^d = Lb^d + r\,(Ub^d - Lb^d) \quad (6)$$

$$(t+1) = \begin{cases} x_{rand}^d + r * \left(x_{rand}^d(t) - x_i^d(t)\right) \\ + \beta * \left(x_{rand}^d(t) - x_i^d(t)\right) i = 1 \\ \\ x_{rand}^d + r * \left(x_{i-1}^d(t) - x_i^d(t)\right) + \beta \\ * \left(x_{rand}^d(t) - x_i^d(t)\right) i = 2, \dots n \end{cases} \quad (7)$$

where, x_{rand}^d is a random position randomly produced in the search space, Ub^d, Lb^d are the upper and lower limits of the d^{th} dimension respectively.

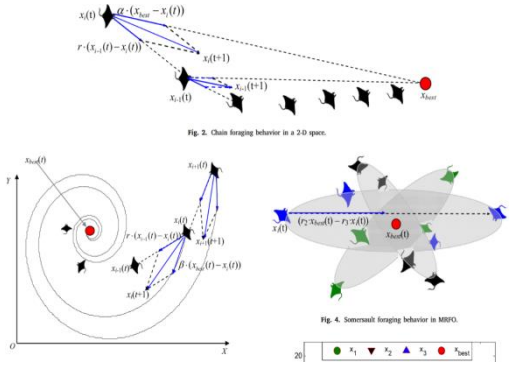

Fig. 2. MRFO (1) Chain (2) Cyclone (3) Somersault behavior in a 3-D space.

- Cyclone foraging: - This behavior treats the food source position as a pivot, causing individuals to swim in a back-and-forth manner while somersaulting to new positions. This results in continuous position updates around the current best solution. The mathematical equation below models this behavior.

$$x_i^d(t+1) = x_i^d(t) + S * \left(r_2 * x_{best}^d - r_3 * x_i^d(t)\right) i = 1, 2, \dots n \quad (8)$$

The somersault factor (S = 2) determines the range of somersaulting for manta rays using random numbers (r1 and r2) in [0, 1]. Eq. (8) allows each individual to move within a new search domain, symmetrically positioned around the best solution found so far. As an individual approaches the optimal solution, the somersault range decreases, guiding all individuals toward convergence. This adaptive reduction occurs as iterations increase. Fig. 4 visually illustrates somersault foraging behavior in MRFO [36].

Like other metaheuristic optimizers, MRFO [36] begins with a random population. In each iteration, individuals update positions relative to neighbors and a reference point. The exploration-exploitation balance is controlled by t/T, decreasing from 1 to t/T. When t/T < rand, the best solution is used for exploitation; otherwise, a random position aids

exploration. Chain and cyclone foraging alternate randomly, while somersault foraging adapts positions to the best solution. This iterative process continues until the stopping criterion is met, yielding the best individual's position and fitness value.

Three individuals undergo 100 iterations in the search space using Eq. (8). Randomly sampled points distribute between current positions and their symmetrical positions around xbest. As distance to xbest decreases, the number of sampled points reduces. Dense points around xbest aid exploitation, while sparse ones promote exploration.

F. Hyperparameter Optimization Process

1) Data preprocessing: After acquiring the dataset, it underwent a series of preprocessing steps involving four key techniques: image resizing, dimensional scaling, data balancing, and data augmentation. In the image processing phase, a two-step approach was utilized to standardize the various image dimensions. Initially, the color space conversion from BGR to RGB was executed to ensure consistent color representation. Subsequently, bicubic interpolation [38] resizing was applied to all images, harmonizing them to a uniform size of (32x32x3) pixels. This process established consistency, laying a solid foundation for subsequent analysis. Addressing the issue of data imbalance [39], the study employed data augmentation techniques encompassing width and height shifting, shearing, rotation, horizontal and vertical flipping, and zooming. These techniques help diversify the dataset and mitigate potential misclassification or overfitting during training and optimization phases [40]. Additionally, the dataset was partitioned into an 80% training subset and a 20% validation subset to facilitate model training and evaluation. This division ensures a robust assessment of model performance and generalizability.

2) Classification and optimization: The optimization process commences with the creation of an initial solution population represented as vectors. Each vector element corresponds to a specific hyperparameter, such as dropout ratio, optimizers, batch size, etc. In this study, 12 well-defined hyperparameters are targeted for optimization, as outlined in Table I. The dimensionality of solution vectors depends on the application of data augmentation.

a) Fitness Function and Hyperparameter Mapping: At the core of the optimization process, the fitness function is utilized to assess the quality of each population solution. This function encompasses three crucial steps:

- Hyperparameter Mapping: To effectively configure the pre-trained transfer learning model, solution elements are mapped to their corresponding actual hyperparameters.
- Top of Form
- Model Creation and Preparation: With mapped hyperparameters, the pre-trained transfer learning CNN model is created and compiled. A diverse set of pre-trained models, including DenseNet121, EfficientB1, EfficientB7, EfficientV2B0, MobileNet, NesNetLarge, ResNet50, VGG16, and VGG19, are harnessed in this study.
- Model Training and Evaluation: Following model creation, the training phase commences, allowing the model to learn from the data. Training progresses for a predetermined number of epochs typically set to 1000 in this study.

b) Balancing Exploration and Exploitation: This approach dynamically traverses the hyperparameter space, striving for a harmonious balance between exploration and exploitation. This equilibrium proves pivotal in uncovering globally optimal solutions. The process iteratively refines the population, gradually converging towards configurations that yield superior model performance.

Fig. 3 depicts the model architecture employed in our optimization process, offering a visual representation of the complex interplay between hyperparameters, fitness evaluation, and model creation.

TABLE I. HYPERPARAMETER SETTINGS

Hyperparameters Settings		
S/noo	**Hyperparameter**	Values
1	**Loss Function**	Binary Crossentropy
2	**Training batch size**	[8, 16,32,64]
3	**Model dropout ratio**	[0- 0.5]
4	**Transfer learning ratio**	0-26 (step = 1)
5	**optimizer**	[Adam(), Nadam(), RMSprop(), Adadelta(), Adagrad(), SGD()]
6	**Rotation Range**	0 - 40 (step = 1)
7	**width shift value.**	[0 - 0.2]
8	**Hight shift value.**	[0 - 0.2]
9	**Shear**	[0 - 0.2]
10	**Zoom**	[0 - 0.2]
11	**Horizontal Flipping**	[Yes, No]
12	**Vertical flipping**	[Yes, No]

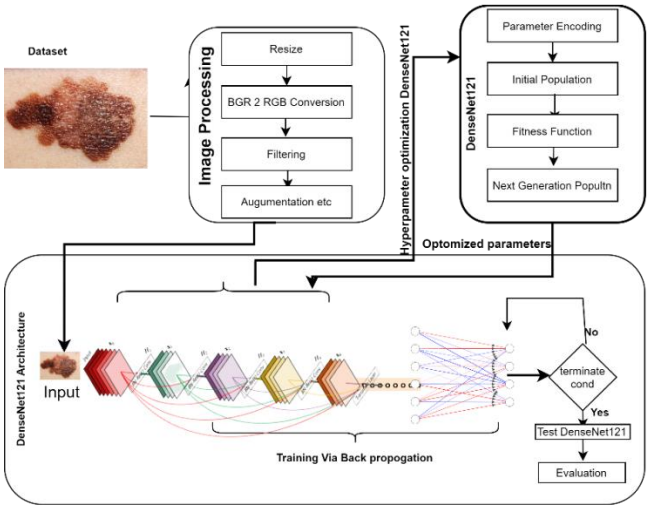

Fig. 3. Proposed model architecture.

G. Evaluation Metrics

The proposed model will undergo evaluation using standard assessment metrics, including Accuracy, Precision, Recall, and F-Score, as mentioned in [38]–[41].

H. Software and Tools

The experiments were conducted in Python, utilizing popular libraries including Keras, Scikit-learn, and OpenCV. The computational environment employed a Dell system equipped with an Intel Xeon CPU, boasting up to 128GB of memory, and a spacious 1TB SSD.

IV. Result and Discussion

In this section, the proposed model was assessed through a comparison with four nature-inspired techniques (GA, WOA, GWO, WHO) and benchmarked against eight state-of-the-art deep learning architectures (EfficientB1, EfficientB7, EfficientV2B0, MobileNet, NesNetLarge, ResNet50, VGG16, and VGG19) on the ISIC 2019 dataset. This dataset comprises 776 skin lesion images encompassing melanoma and non-melanoma classes.

Data preprocessing was undertaken, including augmentation to expand the dataset to 8000 images, and an 80%-20% random split was applied for training and validation purposes. For image resizing, bilinear interpolation was utilized. The experiments were conducted using the Keras library with the hyperparameters detailed in Table I.

The model is comprised of 7.2 million parameters, which includes 121 convolutional layers and three fully connected layers. Training was executed over 1000 epochs, involving 12 distinct hyperparameter configurations and the incorporation of early stopping with a patience of 15.

Proposed model performance across top 10 epochs

Top of Form

In Table II, we observed that the model's performance across the top 10 epochs consistently exhibited high accuracy, ranging from 0.9908 to 0.9926, and AUC, ranging from 0.9898 to 0.9974. Additionally, precision, recall, and the F1-Score consistently demonstrated strong values, indicating proficient classification capabilities with minimal discrepancies. These results underscore the model's robust and stable performance, making it well-suited for accurate for melanoma classification.

TABLE II. Performance of Proposed Model across Top 10 Epochs

Epoch	Accuracy	AUC	Precision	Recall	F1-Score	Loss
38	0.9926	0.9956	0.9406	0.8796	0.9091	0.0217
37	0.9924	0.9974	0.9447	0.8704	0.906	0.0216
34	0.9916	0.9952	0.9529	0.8426	0.8943	0.0241
35	0.9916	0.9951	0.9261	0.8704	0.8974	0.0244
40	0.9914	0.9937	0.9343	0.8565	0.8937	0.0224
33	0.9912	0.9974	0.9430	0.8426	0.89	0.0237
25	0.9910	0.9900	0.9381	0.8426	0.8878	0.0331
32	0.9908	0.9942	0.9378	0.8380	0.8851	0.0259
36	0.9908	0.9898	0.9163	0.8611	0.8878	0.0283

Model Validation Performance Across Top 10 Epochs

In correspondence to Table III, the model consistently demonstrates strong validation performance, with accuracy ranging from 0.9586 to 0.9734 and AUC values between 0.8164 and 0.8588. Validation loss falls in the range of 0.1458 to 0.2014, indicating close alignment with the ground truth. Precision values consistently exceed 0.5, indicating minimal false positives, while recall values vary, indicating the model's ability to correctly identify positive samples. These findings highlight the model's robust and consistent validation performance, ideal for tasks demanding accurate classification.

TABLE III. Validation Performance across Top 10 Epochs

Epoch	Val. Acc	Val AUC	Val Prec.	Val recall	F1-Score	Val Loss
38	**0.9680**	**0.8179**	**0.6857**	**0.4444**	**0.5393**	**0.1980**
37	0.9688	0.8519	0.7059	0.4444	0.5455	0.1521
34	0.9711	0.8164	0.8148	0.4074	0.5432	0.1652
35	0.9586	0.8588	0.5085	0.5556	0.531	0.1660
40	0.9688	0.8258	0.7333	0.4074	0.5238	0.2014
33	0.9695	0.8353	0.7419	0.4259	0.5412	0.1774
25	0.9734	0.8502	0.8333	0.4630	0.5952	0.1458
32	0.9719	0.8396	0.8462	0.4074	0.55	0.1692
36	0.9664	0.8288	0.6774	0.3889	0.4941	0.1653

From Fig. 4, the performance of DenseNet121 model was evaluated under three different scenarios: 100 epochs with a 5000-sample dataset, 500 epochs with an 8000-sample dataset, and 1000 epochs with an 8000-sample dataset. The model with 1000 epochs on the larger dataset demonstrated the highest accuracy (99.26%), AUC (99.56%), and F1-Score (0.9091), indicating its superior ability to classify positive and negative instances. It also achieved the lowest loss (0.0217) and high precision (94.06%) and recall (87.96%) values. It is observed that increasing the number of training epochs and dataset size consistently leads to better model performance across various metrics.

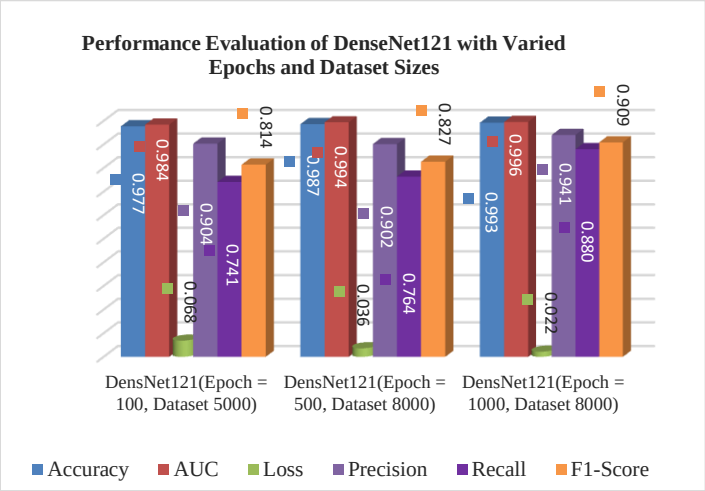

Fig. 4. Comparison with different configuration.

Comparative Analysis of the Proposed Model and Other Architectures

TABLE IV. COMPARATIVE ANALYSIS OF MODEL PERFORMANCE

Model	Accuracy	AUC	F1-Score	Precision	Recall	Loss
Proposed Model	0.9926	0.9956	0.9091	0.9406	0.8796	0.0217
EfficientB1	0.9580	0.4972	0.0092	1.0000	0.0046	0.1815
EfficientB7	0.9232	0.4824	0.1495	0.5000	0.0879	0.6742
EfficientNetV2B0	0.9304	0.5135	0.1946	0.6098	0.1157	0.6694
MobileNet	0.9910	0.9980	0.9012	0.9397	0.8657	0.0221
NesNetLarge	0.6911	0.7096	0.5127	0.6547	0.4213	0.6968
ResNet50	0.9656	0.9134	0.4013	0.7564	0.2731	0.1076
VGG16	0.9865	0.9937	0.8244	0.9152	0.7500	0.0369
VGG19	0.9797	0.9675	0.7263	0.8415	0.6389	0.0628

The presented data showcases the performance evaluation of the proposed model (see Table IV) in comparison to several established DL architectures across various metrics. The proposed model outperforms all other models with an impressive accuracy of 99.26%, an AUC of 99.56%, and a commendable F1-Score of 0.9091. Furthermore, it exhibits a high precision of 94.06% and recall of 87.96%, indicating its robustness in correctly classifying positive instances while minimizing false positives. In contrast, models like EfficientB1, EfficientB7, EfficientNetV2B0, NesNetLarge, ResNet50, VGG16, and VGG19 show lower accuracy and F1-Scores, reflecting their limitations in effectively handling the given task. MobileNet, while achieving a high accuracy and AUC, falls slightly behind the proposed model in F1-Score, precision, and recall. These results underscore the superior performance of the proposed model in comparison to established architectures, demonstrating its potential for melanoma classification. Additionally, the low loss value of 0.0217 for the proposed model further confirms its proficiency in minimizing prediction errors during training.

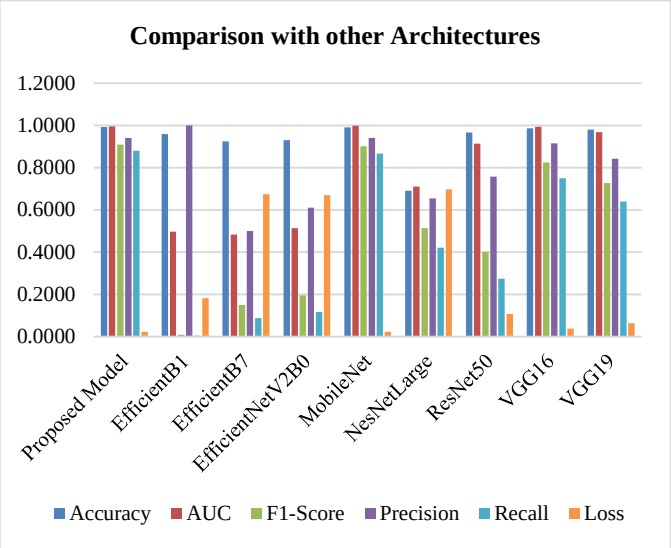

Fig. 5. Comparison with other architectures.

Table V shows the In-Depth Analysis of Proposed DensNet121 Model in Comparison with Prior Literature.

TABLE V. ANALYSIS OF PROPOSED MODEL IN COMPARISON WITH PRIOR LITERATURE

Reference	Model	Metaheuristics	Accuracy	Precision	Recall	F1_Score
[42]	CNN	GWO	98.3			
[43]	CNN	WHO	96.0			
[23]	ALexNet	GA	0.8166	0.8118	0.8167	0.8092
[23]	CNN	PSO	90.01	0.0.898	0.90	0.897
Proposed Model	DensNet121	MRFO	0.9926	0.9406	0.8796	0.9091

In comparison with other literature, the proposed DensNet121-based model optimized with MRFO exhibits superior performance. Fig. 5 shows the comparison with other architectures. It achieves an accuracy of 99.26%, surpassing previous models utilizing different metaheuristics such as GWO (98.3%) and WHO (96%). Moreover, the proposed model's precision (0.9406) and recall (0.8796) outperform an AlexNet model optimized with GA (accuracy: 81.66%, precision: 81.18%, recall: 81.67%). Additionally, the F1-Score of the proposed model (0.9091) exceeds that of a CNN optimized with PSO (accuracy: 90.01%, F1-Score: 0.897). These findings highlight the effectiveness of the proposed model in achieving superior accuracy and a balanced precision-recall trade-off when compared to existing literature.

V. CONCLUSION

In conclusion, the manual classification of melanoma from dermoscopic images presents significant challenges, even for experts, highlighting the pressing need for efficient automation solutions. This paper introduces an automated and cost-effective model built upon the DenseNet121 architecture, harnessed by the MRFO metaheuristic algorithm for melanoma classification. Through careful optimization of critical CNN hyperparameters, this model significantly enhances the architecture's ability to tackle melanoma classification effectively.

The study employs data preprocessing techniques, including bilinear interpolation-based image resizing, and adeptly encodes CNN hyperparameters to facilitate optimization. To validate the model's efficacy, a comprehensive evaluation was conducted, comparing it with hyperparameter optimization techniques GA, PSO, GWO and WHO on the ISIC skin cancer dataset. Additionally, the model was benchmarked against eight state-of-the-art deep learning

architectures (EfficientB1, EfficientB7, EfficientV2B0, MobileNet, NesNetLarge, ResNet50, VGG16, and VGG19).

The final results demonstrated the superiority of our proposed model in terms of accuracy, AUC, precision, F1 Score and loss rates during training, testing, and validation phases, surpassing the performance of other optimization methods explored in this experiment.

As a direction for future research, further refinement and exploration of this model can lead to even more robust and accurate melanoma classification systems, potentially contributing to early diagnosis and improved patient outcomes in the field of dermatology.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to Yayasan Universiti Teknologi Petronas, Malaysia, for their invaluable support of this research through the research grants YUTP-FRG (015LC0-487). Their financial assistance has played a crucial role in the successful completion of this work and has significantly contributed to the advancement of our research objectives.

REFERENCES

[1] N. L. Bolick and A. C. Geller, "Epidemiology of Melanoma," Hematol. Oncol. Clin. North Am., vol. 35, no. 1, pp. 57–72, 2021, doi: 10.1016/j.hoc.2020.08.011.

[2] "Melanoma," 2023. https://www.mayoclinic.org/diseases-conditions/melanoma/symptoms-causes/syc-20374884 (accessed Sep. 09, 2023).

[3] B. N. E. Cancer, "Melanoma: Statistics," 2023. https://www.cancer.net/cancer-types/melanoma/statistics (accessed Sep. 09, 2023).

[4] C. N. E. Board, "Cancer Net," 2023. https://www.cancer.net/cancer-types/skin-cancer-non-melanoma/statistics (accessed May 17, 2023).

[5] World Cancer Research Fund, "Cancer Facts and Figures 2021," World Cancer Research Fund International. pp. 1–4, 2021. [Online]. Available: http://www.wcrf.org/int/cancer-facts-figures/worldwide-data

[6] J. A. Fee, F. P. McGrady, and N. D. Hart, "Dermoscopy use in primary care: a qualitative study with general practitioners," BMC Prim. Care, vol. 23, no. 1, pp. 1–10, 2022, doi: 10.1186/s12875-022-01653-7.

[7] J. Chai, H. Zeng, A. Li, and E. W. T. Ngai, "Deep learning in computer vision: A critical review of emerging techniques and application scenarios," Mach. Learn. with Appl., vol. 6, p. 100134, 2021, doi: https://doi.org/10.1016/j.mlwa.2021.100134.

[8] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," Insights Imaging, vol. 9, no. 4, pp. 611–629, 2018, doi: 10.1007/s13244-018-0639-9.

[9] L. Cai, J. Gao, and D. Zhao, "A review of the application of deep learning in medical image classification and segmentation," Ann. Transl. Med. Vol 8, No 11 (June 15, 2020) Ann. Transl. Med., 2020, [Online]. Available: https://atm.amegroups.org/article/view/36944

[10] I. Sonata, Y. Heryadi, L. Lukas, and A. Wibowo, "Autonomous car using CNN deep learning algorithm," J. Phys. Conf. Ser., vol. 1869, no. 1, 2021, doi: 10.1088/1742-6596/1869/1/012071.

[11] W. Wang and J. Gang, "Application of Convolutional Neural Network in Natural Language Processing," in 2018 International Conference on Information Systems and Computer Aided Education (ICISCAE), 2018, pp. 64–70. doi: 10.1109/ICISCAE.2018.8666928.

[12] L. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," Neurocomputing, vol. 415, pp. 295–316, 2020, doi: https://doi.org/10.1016/j.neucom.2020.07.061.

[13] M. Wojciuk, Z. Swiderska-Chadaj, K. Siwek, and A. Gertych, "The Role of Hyperparameter Optimization in Fine-Tuning of Cnn Models," SSRN Electron. J., 2022, doi: 10.2139/ssrn.4087642.

[14] A. W. Salehi et al., "A Study of CNN and Transfer Learning in Medical Imaging: Advantages, Challenges, Future Scope," Sustainability, vol. 15, no. 7. 2023. doi: 10.3390/su15075930.

[15] A. Gogna and A. Tayal, "Metaheuristics: review and application," J. Exp. \& Theor. Artif. Intell., vol. 25, no. 4, pp. 503–526, 2013, doi: 10.1080/0952813X.2013.782347.

[16] R. Chen, X. Tang, and X. Li, "Adaptive Stochastic Gradient Descent Method for Convex and Non-Convex Optimization," Fractal and Fractional, vol. 6, no. 12. 2022. doi: 10.3390/fractalfract6120709.

[17] M. Abdel-Basset, L. Abdel-Fatah, and A. K. Sangaiah, Metaheuristic algorithms: A comprehensive review. Elsevier Inc., 2018. doi: 10.1016/B978-0-12-813314-9.00010-4.

[18] Y. Kumar, A. Koul, R. Singla, and M. F. Ijaz, "Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda.," J. Ambient Intell. Humaniz. Comput., vol. 14, no. 7, pp. 8459–8486, 2023, doi: 10.1007/s12652-021-03612-z.

[19] A. K. Adepu, S. Sahayam, U. Jayaraman, and R. Arramraju, "Melanoma classification from dermatoscopy images using knowledge distillation for highly imbalanced data," Comput. Biol. Med., vol. 154, no. July 2022, p. 106571, 2023, doi: 10.1016/j.compbiomed.2023.106571.

[20] E. Yang et al., "Machine learning modeling and prognostic value analysis of invasion-related genes in cutaneous melanoma," Comput. Biol. Med., vol. 162, no. March, p. 107089, 2023, doi: 10.1016/j.compbiomed.2023.107089.

[21] H. K. Gajera, D. R. Nayak, and M. A. Zaveri, "A comprehensive analysis of dermoscopy images for melanoma detection via deep CNN features," Biomed. Signal Process. Control, vol. 79, no. P2, p. 104186, 2023, doi: 10.1016/j.bspc.2022.104186.

[22] J. Daghrir, L. Tlig, M. Bouchouicha, and M. Sayadi, "Melanoma skin cancer detection using deep learning and classical machine learning techniques: A hybrid approach," 2020 Int. Conf. Adv. Technol. Signal Image Process. ATSIP 2020, no. C, pp. 3–7, 2020, doi: 10.1109/ATSIP49331.2020.9231544.

[23] F. Golnoori, F. Z. Boroujeni, and A. Monadjemi, "Metaheuristic algorithm based hyper-parameters optimization for skin lesion classification," Multimed. Tools Appl., vol. 82, no. 17, pp. 25677–25709, 2023, doi: 10.1007/s11042-023-14429-7.

[24] M. Rahman, M. K. Nasir, and S. I. Khan, "Hybrid Feature Fusion and Machine Learning Approaches for Melanoma Skin Cancer Detection," no. January, 2022, doi: 10.20944/preprints2.

[25] N. Nida, A. Irtaza, A. Javed, M. H. Yousaf, and M. T. Mahmood, "Melanoma lesion detection and segmentation using deep region based convolutional neural network and fuzzy C-means clustering," Int. J. Med. Inform., vol. 124, no. January, pp. 37–48, 2019, doi: 10.1016/j.ijmedinf.2019.01.005.

[26] E. Pérez and S. Ventura, "A framework to build accurate Convolutional Neural Network models for melanoma diagnosis," Knowledge-Based Syst., vol. 260, p. 110157, 2023, doi: 10.1016/j.knosys.2022.110157.

[27] A. K. Adepu, S. Sahayam, U. Jayaraman, and R. Arramraju, "Melanoma classification from dermatoscopy images using knowledge distillation for highly imbalanced data," Comput. Biol. Med., vol. 154, no. July 2022, p. 106571, 2023, doi: 10.1016/j.compbiomed.2023.106571.

[28] S. Salem Ghahfarrokhi, H. Khodadadi, H. Ghadiri, and F. Fattahi, "Malignant melanoma diagnosis applying a machine learning method based on the combination of nonlinear and texture features," Biomed. Signal Process. Control, vol. 80, no. P1, p. 104300, 2023, doi: 10.1016/j.bspc.2022.104300.

[29] R. Indraswari, R. Rokhana, and W. Herulambang, "Melanoma image classification based on MobileNetV2 network," Procedia Comput. Sci., vol. 197, pp. 198–207, 2021, doi: 10.1016/j.procs.2021.12.132.

[30] R. Patil and S. Bellary, "Machine learning approach in melanoma cancer stage detection," J. King Saud Univ. - Comput. Inf. Sci., vol. 34, no. 6, pp. 3285–3293, 2022, doi: 10.1016/j.jksuci.2020.09.002.

[31] Ç. Suiçmez, H. Tolga Kahraman, A. Suiçmez, C. Yılmaz, and F. Balcı, "Detection of melanoma with hybrid learning method by removing hair from dermoscopic images using image processing techniques and wavelet transform," Biomed. Signal Process. Control, vol. 84, no. July 2022, 2023, doi: 10.1016/j.bspc.2023.104729.

[32] Y. Nancy Jane, S. K. Charanya, M. Amsaprabhaa, P. Jayashanker, and K. Nehemiah H., "2-HDCNN: A two-tier hybrid dual convolution neural network feature fusion approach for diagnosing malignant melanoma," Comput. Biol. Med., vol. 152, no. November 2022, p. 106333, 2023, doi: 10.1016/j.compbiomed.2022.106333.

[33] "ISIC2019." https://challenge.isic-archive.com/data/#2019 (accessed Feb. 04, 2023).

[34] N. Hasan, Y. Bao, A. Shawon, and Y. Huang, "DenseNet Convolutional Neural Networks Application for Predicting COVID-19 Using CT Image," SN Comput. Sci., vol. 2, no. 5, p. 389, 2021, doi: 10.1007/s42979-021-00782-7.

[35] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, vol. 2017-Janua, pp. 2261–2269, 2017, doi: 10.1109/CVPR.2017.243.

[36] W. Zhao, Z. Zhang, and L. Wang, "Manta ray foraging optimization: An effective bio-inspired optimizer for engineering applications," Eng. Appl. Artif. Intell., vol. 87, no. September 2019, p. 103300, 2020, doi: 10.1016/j.engappai.2019.103300.

[37] S. Mirjalili and A. Lewis, "The Whale Optimization Algorithm," Adv. Eng. Softw., vol. 95, pp. 51–67, 2016, doi: 10.1016/j.advengsoft.2016.01.008.

[38] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer Statistics, 2021," CA. Cancer J. Clin., vol. 71, no. 1, pp. 7–33, 2021, doi: 10.3322/caac.21654.

[39] C. Szegedy et al., "Going Deeper with Convolutions," pp. 1–9, 2015.

[40] V. Badrinarayanan, A. Kendall, R. Cipolla, and S. Member, "SegNet : A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," vol. 39, no. 12, pp. 2481–2495, 2017.

[41] Y. Li, J. Xiao, Y. Chen, and L. Jiao, "Neurocomputing Evolving deep convolutional neural networks by quantum b ehave d particle swarm optimization with binary encoding for image classification," vol. 362, pp. 156–165, 2019, doi: 10.1016/j.neucom.2019.07.026.

[42] R. Mohakud and R. Dash, "Designing a grey wolf optimization based hyper-parameter optimized convolutional neural network classifier for skin cancer detection," vol. 34, pp. 6280–6291, 2022.

[43] B. Zhou and B. Arandian, "An Improved CNN Architecture to Diagnose Skin Cancer in Dermoscopic Images Based on Wildebeest Herd Optimization Algorithm," Comput. Intell. Neurosci., vol. 2021, 2021, doi: 10.1155/2021/7567870.