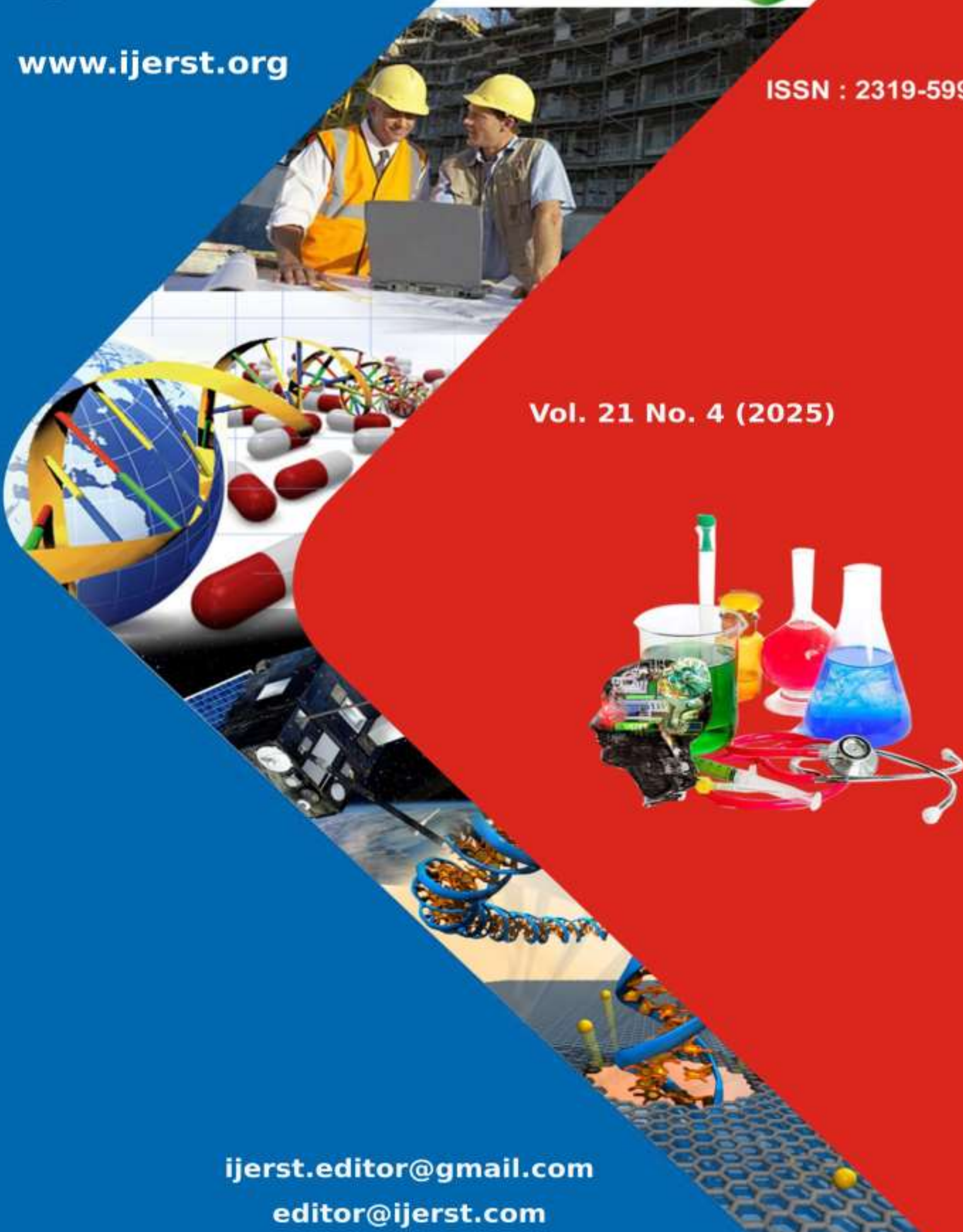
International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 21 No. 4 (2025)



ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com
Vol. 21, No. 4, 2025

Research Paper

HEALTH DATA ANALYSIS FOR CANCER PREDICTION BY OPTIMIZING FEATURES & DEEP LEARNING MODEL

Venkata Karthik Macharla
Department of Industrial Management, D2SOL
The University of Texas at Tyler, United States

Abstract: Skin cancer is a severe form of cancer that, if left untreated, can lead to fatal consequences. Detecting it at an early stage enables more effective treatment and helps stop the disease from advancing. This paper has proposed a model that classify the health data image of skin cancer. Input image was optimized by the artificial immune optimization algorithm. Optimized image was used for the feature extraction. Extracted features were used for the training of mathematical model. Experiment was done on real dataset images of skin cancer. Result shows that proposed HIDADL (Health Image Data Analysis using Deep Learning) model has increases the correct class detection accuracy.

Index Terms— Digital Image Processing, Skin Cancer Diagnosis, Machine learning, Medical Image Diagnosis, Feature Extraction, Segmentation.

Received: 24-09-2025 Accepted: 28-10-2025 Published: 04-11-2025

I. INTRODUCTION

Cancer refers to the uncontrolled growth and division of tissues within a specific part of the body [1]. Among the rapidly increasing types of cancers worldwide, skin cancer has become one of the most prevalent. It occurs when abnormal skin cells proliferate uncontrollably, leading to tumor formation and potential spread if untreated [2]. In modern healthcare, Computer-Aided Diagnostic (CAD) systems play a crucial role in analyzing and interpreting medical images for accurate and timely diagnosis [1]. CAD is particularly significant in diagnostic radiography and medical imaging research, as its proper application can facilitate the early identification of diseases. Such early detection directly contributes to timely treatment interventions, which can significantly improve patient survival rates [2]. For instance, the chances of effective treatment outcomes are closely tied to the ability to detect cancers in their early stages [3].

Cancer remains one of the leading causes of human mortality worldwide, with skin cancer accounting for a considerable proportion of cases. Typically, it develops in areas of the skin exposed to sunlight for prolonged periods, though it can also occur in less exposed regions. Because it originates in the epidermis—the outermost layer of the skin—skin cancer is relatively visible compared to cancers that develop deeper within the body. This visibility makes CAD-based analysis of skin lesion images particularly valuable, as it allows for preliminary diagnosis without requiring additional clinical information [4].

Early diagnosis is essential in reducing the morbidity and mortality associated with skin cancer by preventing its progression to advanced stages [3]. Despite this, large-scale population-wide screening programs have not yet proven to be cost-effective or successful in reducing mortality rates. Instead, personalized screening programs targeting high-risk individuals are considered a more viable strategy to address the ongoing rise in skin cancer cases [5].

In recent years, deep learning techniques have transformed skin cancer diagnostics, offering enhanced accuracy in image classification and lesion detection. However, challenges remain due to artifacts such as hair, light reflections, and air bubbles that complicate skin lesion segmentation [5]. Another major limitation is the scarcity of high-quality training datasets in the medical domain, which hinders the performance of machine learning models. A widely adopted solution to this issue is data augmentation, which artificially increases dataset variability by introducing transformations that help models generalize better. While significant efforts have been directed toward designing and refining advanced machine learning architectures, relatively fewer studies have focused on improving augmentation techniques, despite their critical role in boosting diagnostic accuracy [6].

II. Related Work

In [7] a hybrid method combining optimization algorithms with Support Vector Machines SVM was proposed for skin cancer detection SCD The optimization algorithms were employed to select the most relevant features while the SVM classifier was then applied to categorize the selected features effectively

In [8] deep learning methods were applied for melanoma diagnosis A Convolutional Neural Network CNN was first employed to extract features which were then used as inputs for two neural network NN models The first NN also a CNN identified the target regions while the second a Recurrent network determined the exact location of the lesion Based on this localization lesion segmentation was ultimately done by the fuzzy K means model

In [9] CNNs were utilized to perform SCD and lesion localization The proposed framework began with preprocessing the input images followed by segmentation using a UNet network The segmented lesion area was then cropped and provided as input to a CNN model which performed the final classification

In [10] an enhanced CNN model derived from the VGG 16 architecture was applied for skin cancer diagnosis The original VGG configuration was modified to better address the challenges of SCD yielding improved accuracy compared to the standard model The improvements included alterations in filter dimensions as well as changes to neural network activation functions

In [11] a deep learning approach based on the MobileNetV2 architecture was developed optimized through a memetic algorithm for hyperparameter tuning This optimization strategy combined both global and local search techniques to optimized essential parameters such as learning rate batch size and number of epochs The resulting model not only achieved high accuracy but also maintained computational efficiency making it practical for real world clinical applications

In [12] the authors highlighted the rising global incidence of skin cancer and the importance of early detection Using the HAM10000 dataset containing 10 015 dermoscopic images across seven skin lesion categories they applied preprocessing methods including sampling dull razor removal and autoencoder based segmentation By employing transfer learning with DenseNet169 and ResNet50 the study demonstrated enhanced accuracy in skin cancer classification

In [13] deep attention mechanisms AMs were explored in the context of medical image classification due to their strong performance when integrated with deep learning DL models Unlike conventional methods emphasize areas of higher diagnostic relevance improving classification outcomes The study specifically examined three attention mechanisms—soft attention hard attention and self attention—each employing a unique strategy to focus on critical regions of the input data

Tembhurne et al [14] proposed an ensemble based framework that combined machine learning feature extraction methods with deep learning models for distinguishing benign from malignant skin cancers Techniques were used for feature computation which were then used to train deep learning models for classification

III. Proposed Methodology

This section presents the proposed HIDADL (Health Image Data Analysis using Deep Learning) framework, illustrated through the block diagram in Fig. 3. The entire workflow is divided into two primary modules. The first module focuses on image pre-processing and optimization of the input images, accomplished using a artificial immune Optimization (AIO) algorithm. The second module performs feature extraction and mathematical model training for classification.

Module 1: Image Pre-Processing & Optimizing

The raw input image is first converted into a two-dimensional matrix, where pixel intensity values range between 0 and 255. To prepare the image for subsequent operations, it is reshaped to facilitate matrix computations. A filtering process is then applied to suppress noise [15], and eliminate unwanted data. Fig. 2 demonstrates an example of the original input image, while Fig. 3 illustrates the result after the pre-processing stage.

Fig. 1 Input image.

Fig. 2 Image after pre-processing operation.

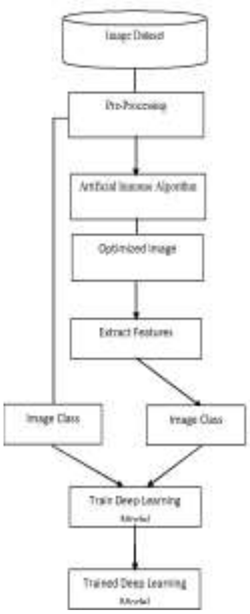


Fig. 3 Block diagram of proposed model.

Artificial Immune Optimization

In this method, each potential solution is considered a antibody, and its affinity represents its fitness. A higher affinity indicates a good-quality solution, while a lower affinity signifies a poor solution in terms of available resources. Flow of whole optimization model is shown in fig. 4.

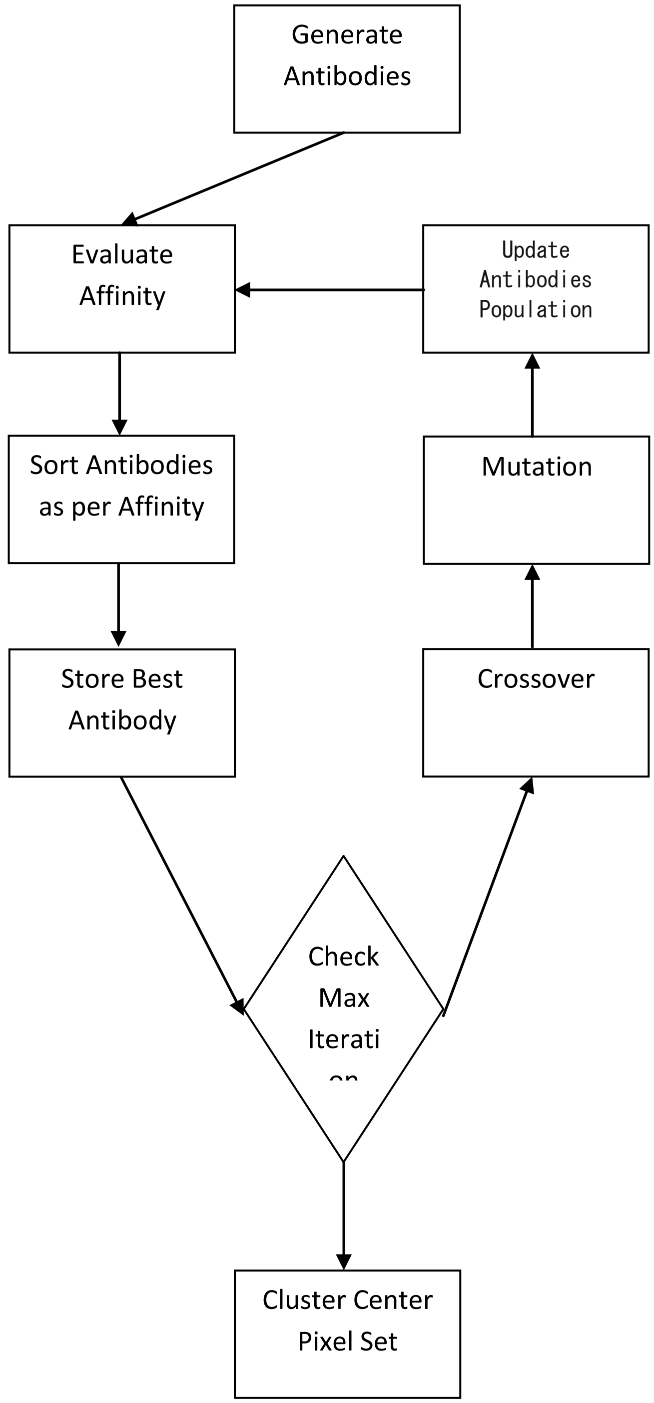


Fig. 4 Block diagram of artificial immune genetic algorithm.

Generate Antibodies: In the initialization phase, a set of potential solutions (antibody) is generated [16]. Each antibody corresponds to a set of possible cluster centers, represented as:

$$H \leftarrow Antibody(s, h)$$

Where s is number of antibody and h is number of cluster center.

Affinity: The fitness of each antibody is determined by measuring the learning accuracy of model as per selected pixels in each cluster. The antibody with the high accuracy is considered the most optimal solution:

$$F_h = Train\_Test(Cluster(_{x,y=1}^{row,col}(Min(H(h,s) - I(x,y)))))$$

where F_h is the fitness value of the hth chromosome, $H_{h,s}$ represents the pixel in the h-th chromosome and s-th cluster, and I(x,y) is the intensity value at pixel position (x , y) x is row count and y is column count in image matrix.

R = Rank(F_h, H)----Eq. 5

Crossover

Pixel cluster center interchange between antibodies is modeled through crossover operations. The process depends on affinity, where new solutions are generated by exchanging features between antibodies [17]. A best antibody havig highest affinity values do crossover operation with other antibodies.

Mutation

To enhance solution diversity, mutation is performed after crossover. Mutation is performed randomly between by replacing the current pixel center value by other.

Update Population After each iterations antibodies having low affinity value is removed from the population. This algorithm ends after max iteration count. Best affinity value antibody is consider as final pixel cluster center set. Through this population optimization process, the pre-processed skin lesion images are enhanced and optimized for the second module of the framework, where feature extraction and deep learning-based classification are performed.

Module II: Extract Histogram & DWT Feature:

From each segmented skin lesion image, two sets of features were derived: histogram features and Discrete Wavelet Transform (DWT) features [18]. Specifically, sixteen values were extracted from the histogram, representing the distribution of pixel intensities, while an additional sixteen values were obtained from the low-frequency sub-band of the DWT, which captures the essential structural information of the image. Together, these features form a compact yet discriminative representation of the input image.

Training of Deep Neural Networks (DNNs)

Instead of classifying pixels in isolation, the proposed framework employs Deep Neural Networks (DNNs) that process the entire input image in a single forward pass to generate a segmented output [19]. Within this architecture, convolutional and pooling layers are applied iteratively, preserving the spatial arrangement of the image while simultaneously extracting high-level representations. This allows the network to retain contextual relationships between pixels, which is critical for accurate medical image classification.

The extracted feature block (EF), after undergoing convolutional transformations, is fed into the DNN as input. The corresponding ground truth label—indicating whether the region contains a tumor or not—is used as the target output during training. Once trained, the DNN is capable of directly processing optimized image blocks and classifying them as tumor or non-tumor. This design provides an efficient and reliable approach for skin cancer image segmentation and classification.

VI. Experiment and Results

Experimental work of proposed work was done on MATLAB software having machine configuration of 4GB RAM, windows operating system and I3 processor. Results of p proposed HIDADL model was compared with existing algorithm Deepskin proposed in [12]. Experimental dataset was taken from [20].

Results

Table 1 Skin cancer image class prediction models comparison based on precision values.

Skin Cancer Dataset	DeepSkin [12]	HIDADL
40	0.7143	0.95
80	0.6512	0.8537
120	0.7087	0.9153
160	0.6448	0.8742
200	0.6364	0.675

Table 1 shows precision values for prediction of skin cancer image class. With increase in testing dataset size prediction accuracy class values get decreases. It was found that use of artificial immune genetic algorithm has increases the class detection precision by 21.38% as compared to Deepskin model proposed in [12].

Table 2 Skin cancer image class prediction models comparison based on recall values.

Skin Cancer Dataset	DeepSkin [12]	HIDADL
40	0.9375	0.95
80	0.9032	0.8333
120	0.8654	0.9153
160	0.8613	0.8688
200	0.6585	0.7352

Recall values of skin cancer image class prediction models were shown in table 2. It was found that proposed model has improved the recall value by 1.78%. Use of Deep learning model convolutional neural network has increased the detection class.

Table 3 Skin cancer image class prediction models comparison based on f-measure values.

Skin Cancer Dataset	DeepSkin [12]	HIDADL
40	0.8293	0.95
80	0.7857	0.8434
120	0.7918	0.9153
160	0.7551	0.8715
200	0.6667	0.6982

Table 3 shows the f-measure values of the comparing models. Further it was found that extraction of histogram color feature after artificial immune algorithm has increased the work performance.

Table 4 Skin cancer image class prediction models comparison based on accuracy values.

Skin Cancer Dataset	DeepSkin [12]	HIDADL
40	82.9268	95.2381
80	78.5714	84.5238
120	79.1837	91.5966
160	75.5102	87.2671
200	66.9388	69.8189

Correct class detection accuracy of the models of image skin cancer prediction was shown in fig. 4. It was found that detection accuracy decreases with number of images in during testing. Detection accuracy was increased by 10.57% as compared to Deepskin model in [12].

V. CONCLUSION

This paper has developed a model that identifies the correct class of skin cancer images. Further it was found that use of image optimization artificial immune genetic algorithm has increases the input data quality. Further visual color histogram feature used in the work has increased the reduces the learning complexity. Extracted features were used for the training of deep learning model has increased the detection accuracy. Experiment was done on real skin cancer image dataset. Result shows that proposed model increases the precision by 21.38% and accuracy by 10.57% as compared to Deepskin model proposed in [12]. In future scholar can adopt the same work for different set of image classification of health dataset.

References

[1]. World Health Organization . Global Health Observatory. World Health Organization; Geneva, Switzerland: 2022.

[2]. Han H.S., Choi K.Y. Advances in nanomaterial-mediated photothermal cancer therapies: Toward clinical applications. Biomedicines. 2021;9:305.

[3]. Dildar M., Akram S., Mahmood A.R., Mahnashi M.H., Alsaiari S.A., Irfan M., Khan H.U., Saeed A.H.M., Ramzan M., Alraddadi M.O. Skin cancer

detection: A review using deep learning techniques. Int. J. Environ. Res. Public Health. 2021;18:5479.

[4]. Hussain SToscano E(2025)Optimized Deep Learning for Mammography: Augmentation and Tailored ArchitecturesInformation10.3390/info1605035916:5(359)

[5]. S. Feng, H. Zhao, F. Shi, X. Cheng, M. Wang, Y. Ma, D. Xiang, W. Zhu, X. ChenCPFNet: context pyramid fusion network for medical image segmentation IEEE Trans. Med. Imag., 39 (10) (2020), pp. 3008-3018.

[6]. X. Jiang, Z. Ge. Data augmentation classifier for imbalanced fault classification IEEE Trans. Autom. Sci. Eng., 18 (3) (2020), pp. 1206-1217.

[7]. Bi, D., Zhu, D., Sheykhahmad, F. R. & Qiao, M. Computer-aided skin cancer diagnosis based on a New meta-heuristic algorithm combined with support vector method. Biomed. Signal Process. Control 68, 102631 (2021).

[8]. Nawaz, M. et al. Skin cancer detection from dermoscopic images using deep learning and fuzzy k-means clustering. Microsc. Res. Tech. 85(1), 339–351 (2022).

[9]. Reis, H. C., Turk, V., Khoshelham, K. & Kaya, S. InSiNet: A deep convolutional approach to skin cancer detection and segmentation. Med. Biol. Eng. Comput. 2022, 1–20 (2022).

[10]. Tabrizchi, H., Parvizpour, S. & Razmara, J. An improved VGG model for skin cancer detection. Neural Process. Lett. 2022, 1–18 (2022).

[11]. Öznacar, T., Varol Kayapunar, N. Advanced skin cancer prediction with medical image data using MobileNetV2 deep learning and optimized techniques. Sci Rep 15, 28962 (2025).

[12]. H. L. Gururaj, N. Manju, A. Nagarjun, V. N. M. Aradhya and F. Flammini, "DeepSkin: A Deep Learning Approach for Skin Cancer Classification," in IEEE Access, vol. 11, pp. 50205-50214, 2023.

[13]. Alotaibi A, AlSaeed D. Skin Cancer Detection Using Transfer Learning and Deep Attention Mechanisms. Diagnostics (Basel). 2025 Jan 3;15(1):99.

[14]. Tembhurne JV, Hebbar N, Patil HY, Diwan T. Skin cancer detection using ensemble of machine learning and deep learning techniques. Multimed Tools Appl. 2023;82(18):27501–24.

[15]. K. Ahmad, J. Khan and M. S. U. D. Iqbal, "A comparative study of Different Denoising Techniques in Digital Image Processing," 2019 8th International Conference on Modeling Simulation and Applied Optimization (ICMSAO), Manama, Bahrain, 2019, pp. 1-6.

[16]. M. W. Yaw, K. H. Chong and K. Kamil, "Transform of Artificial Immune System algorithm optimization based on mathematical test function," 2016 6th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), Penang, Malaysia, 2016, pp. 147-150

[17]. Oday A. Ahmed, K.H. Chong, S.P. Koh, Chong Tak Yaw, Jagadeesh Pasupuleti, Artificial immune systems (GA-AIS) enabled power loss mitigation in distributed generation: X3PAIS optimization approaches, Heliyon, Volume 10, Issue 18, 2024.

[18]. Martey EM, Lei H, Li X, Appiah O. Image Representation Using Stacked Colour Histogram. Algorithms. 2021; 14(8):228.

[19]. Yao Liu, Hongbin Pu, Da-Wen Sun, Efficient extraction of deep image features using convolutional neural network (CNN) for applications in detecting and analysing complex food matrices, Trends in Food Science & Technology, Volume 113, 2021, Pages 193-204.

[20]. http://www.dermnet.com/dermatology-pictures-skin-disease-pictures