Melanoma Detection Based on SVM Using MATLAB

Radhwan M. W. Khaleel* and Nasseer M. Basheer

Department of Medical Instrumentation Techniques Engineering, Technical Engineering College Mosul, Northern Technical University, Mosul, Iraq
*Correspondence: rudwanalallaf@gmail.com (R.M.W.K.)

***Abstract*—Skin cancer has become the fifth-most dangerous type of cancer. Melanoma, the most ferocious type of skin cancer, should be detected and treated to reduce the risk of spreading to the rest of the body's organs. This study aims to provide fast and painless detection of skin cancer using image processing, including enhancement and extraction of interesting features for the characterization and classification of infected skin images into melanoma or non-melanoma in MATLAB. The features used for texture analysis of inserted images are the Gray Level Co-occurrence Matrix (GLCM) and Local Binary Pattern (LBP). The classification of melanoma and non-melanoma is done by training a Support Vector Machine (SVM) using the radial basis function kernel. The accuracy of testing is 94.87%.**

***Keywords*—Melanoma, Support Vector Machine (SVM), Radial Bases Function (RBF), image processing, Local Binary Pattern (LBP), Gray Level Co-occurrence Matrix (GLCM), MATLAB**

I. Introduction

Skin is the outer organ in the human body, it is responsible for protecting the internal organs and constitutes 16% of the body weight. The skin may be exposed to several injuries or diseases, one of these diseases is skin cancer. One of the most important causes of skin cancer is exposure to Ultra violet (UV) from the sun [1].

Skin cancer is divided into three basic types: Basal Cell Carcinoma (BCC), which is the most common type, represents 80% of the injuries, and is the least dangerous. The second type is Squamous Cell Carcinoma (SCC), which is less common than the first type and it represents almost 16–17% of skin cancers around the world. The third type is melanoma, which is the most dangerous type but the least prevalent, as it represents 4% of skin cancers and is responsible for the deaths of 75% of skin cancer patients. Melanoma can be cured if detected and treated in the early stages, but if it is detected too late, it may spread to other parts of the body [2].

There are several ways to detect skin cancer. One of these methods is the noninvasive and painless detection of skin cancer using a Dermatoscope. This is a device used to magnify skin lesions and increase the accuracy of detection [3].

The Support Vector Machine (SVM) is one of the machine learning algorithms that achieved good results because it has a good and strong mathematical basis. It can be used in classification for skin cancer as a good, quick and painless method [4].

A series of pre-processing operations, including morphological operations, are used to improve and extract information from input images. This information is then entered into the SVM algorithm to train it and enable it to understand and analyze the data for carrying out the diagnostic task. The data used in this study are images taken by a Dermatoscope device. This dataset is available for research purposes. It contained melanoma, other types of skin cancer, and skin diseases which are non-melanoma. The diagnostic results obtained have an accuracy of 94.87%.

II. Literature Review

Nasr-Esfahani *et al.* [5] proposed Convolutional Neural Network (CNN) as a method to diagnose melanoma. The number of dataset images used was 170 images taken by a digital camera (not by Dermatoscope), then increased to 6120 images by rotating, cropping, and resizing the original images. Then divide it randomly into 80% for training and 20% for testing. The accuracy of the result was relatively low (81%).

Hasan *et al.* [6] suggested CNN for feature extraction and classification. The dataset contained 23,907 dermoscopic images taken from the International Skin Imaging Collaboration (ISIC) dermoscopic archive. The dataset was divided into two classes, one of which contains all the cancer dermoscopic images and the other of which contains non-cancerous dermoscopic images. The accuracy for the test was 89.5%, while the training accuracy was 93.7%.

Murugan *et al.* [7] proposed three methods to classify and diagnose skin cancer: SVM, Random Forest, and KNN (K-Nearest Neighbor). They compared the accuracy results obtained depending on the features extracted from the images, which were the shape (Asymmetry, Border, Color and Diameter) of the ABCD rule and (Gray Level Co-occurrence Matrix) GLCM. The dataset used in this paper contained 1000 images. The result of comparisons gave SVM the best accuracy with an accuracy of 89.43%

Manuscript received June 9, 2023; revised July 4, 2023; accepted July 26, 2023.

when using the ABCD rule for features extraction, while when using shape feature extraction, it was 82.31%. Using GLCM feature extraction, it was 85.72%. These metrics are considered low according to other methods. Also, they used a median filter to reduce noise such as hair and bubbles in the image.

Thaajwer and Ishanka [8] designed computer-aided detection system for melanoma early detection, using a support vector machine classifier depending on GLCM, shape, and color feature extraction. The dataset that was used had 600 images. The accuracy of the results was 83%.

Babu and Peter [9] suggested Radial Bases Function (RBF) based on an SVM classifier to diagnose skin cancer depending on Histogram of Oriented Gradients (HOG) feature extraction from images and using a median filter to reduce the noise. The dataset used was that available from the International Skin Imaging Collaboration 2018 (ISIC 2018), containing 10,015 images. The accuracy of the results was 76%.

In 2022, Alsarraf and Ucan [10] proposed a method for analyzing the skin data images for enhanced melanoma detection. This method included four phases: first, segmentation using K-means clustering and thresholding. Second, using the Sobel operator median filter in pre-processing. Third, in feature extraction, grey-level co-occurrence matrices and local binary patterns were used. Finally, using a support vector machine for classification. The data used ISIC dermoscopic images. The accuracy of this method was 93.28%. They suggested sending the clinical images to a dermatologist via a mobile application (such as WhatsApp) for examination.

In 2022, Sharma and Chandrahaas [11] suggested an approach for skin disease diagnosis. The data used 1285 images, including 120 images for testing. Features are extracted using the Histogram of Oriented Gradients (HoG). Naïve Bayes and SVM algorithms were used for classification, and they compared the results. Accuracy was 77.23% when Naïve Bayes was used and 83% when SVM was used.

In 2023, Dimililer and Sekeroglu [12] designed a transfer learning model using a convolutional neural network (CNN) for skin lesion image classification. In the suggested method, pre-processing and data augmentation were not used. The PAD-UFES-20 dataset contains 2298 images taken by a smartphone camera. The accuracy was 86% in detecting cancerous lesions.

It is clear from the above that it is good to use the SVM Classifier because it can give good results in diagnosing skin cancer, especially when using effective feature extraction.

The main contribution of this study can be summarised as follows: unify the size of the used database images, as the original images in the database were of different sizes and aspect ratios. Also, the use of a combination of algorithms to extract features gave more accurate data for the image details, through which the system reached a good accuracy for melanoma diagnosis.

A thorough explanation of the procedures follows a mention of the study's theoretical foundation. More details and explanations about the database used.

Discussion of the obtained results, and finally, presenting the conclusions reached by this study.

III. Theoretical Background

In this work, the focus is on Innovating melanoma detection technology, following specific steps as shown in Fig. 1.

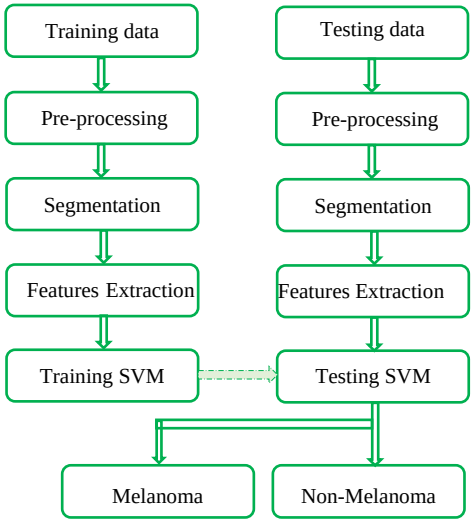


Figure 1. Block diagram of proposed system.

A. Input Data

The database used in this work contains Dermatoscope images of different sizes that have been unified using digital image processing.

- *Unsharp masking:* this is a high-pass filter that is highly efficient in detecting the edges, showing the fine details in the image, and improving the contrast along these edges to produce a sharper image [13].
- *Bicubic interpolation:* is a complex technique used for image resizing; it produces the best results for soft details. It uses a 4×4 grid of adjacent pixels to estimate a pixel's value in the new image. The interpolation value is calculated using Eq. (1), as in [13]:

$$v(x, y) = \sum_{i}^{3} \sum_{j=0}^{3} a_{ij} x^{i} y^{j} \qquad (1)$$

where $v(x, y)$ is interpolation value at (x, y). The coefficients a_{ij} are calculated using the sixteen closest neighbors of the point $(x, y)$to solve the sixteen equations with sixteen unknowns.

B. Pre-processing

- *Grayscaled image:* To simplify calculations and reduce computational burden, images used in this step are converted to grayscale, which consists of a single channel, unlike the color image, which has three (red, green, and blue) channels.
- *Histogram equalization:* expands the dynamic range of the image to produce a more uniformly distributed brightness scale [14].

C. Segmentation

- *Morphological operations:* use small matrices called structuring elements to modify image pixels [13]. Erosion makes objects smaller and thinner, while dilation makes them larger. The opening procedure removes small objects, smooths larger object edges, and separates connected objects. The closing procedure preserves outer object borders while closing gaps and holes and can combine separated objects that are close. These operations are useful in image segmentation to isolate and extract specific objects or features in an image. Note that the amount of erosion or dilation of the image is determined by the size and shape of the structural element, not the original image [15].
- *Thresholding:* The fundamental concept of segmentation using intensity thresholding is very important. By selecting a threshold value so that pixels with values above the threshold are assigned to one region and those with values below the threshold are assigned to another. According to the straightforward requirement, thresholding converts an intensity image $bw(x, y)$ into a binary image $I(x, y)$, T is the threshold [16].

$$bw(x,y) = \begin{cases} 1, & \text{if } I(x,y) > T \\ 0, & \text{Otherwise} \end{cases} \quad (2)$$

D. Features Extraction

Feature extraction includes statistical-based features and shape features.

1) Statistical based features:

- *Local Binary Pattern (LBP):* It is a technique for analyzing textures that describes an image's local structure. LBP operates by comparing each pixel's intensity value with the intensity values of its 3×3 surrounding pixels. In particular, the LBP operator assigns each pixel in the image a binary code by comparing the intensity of the central pixel with the intensity values of its neighbors. A pixel is given a value of 1 if the intensity of the center pixel is higher than or equal to the intensity of a neighboring pixel; otherwise, it is given a value of 0 [17].
- *Gray Level Co-occurrence Matrix (GLCM):* is used to analyze image texture. It describes the probability of two pairs of gray levels occurring together in an image, allowing for the extraction of texture features to differentiate between various textures. This study used GLCM to extract four important different features [18], contrast, correlation, energy, and homogeneity.
- *Skewness:* measures asymmetry in the probability distribution over the mean of a real-valued random variable, with values being positive, negative, or undefinable [19].
- *Kurtosis:* on the other hand, it describes the shape of a probability distribution for a real-value random variable [19].

2) Shape features:

- *Diameter:* is a significant feature for detecting skin cancer. Lesions with a diameter greater than 6mm are indicative of melanoma, while those with a smaller diameter are non-melanoma [20].

E. Classification

Support Vector Machines (SVM) are a model for supervised learning that uses learning algorithms to analyze data for classification and regression purposes. SVM is specifically designed for data with two classes and identifies the best hyperplane to separate the data points of each class, with the ideal hyperplane being the one with the largest margin between the classes. The support vectors refer to the data points closest to the separating hyperplane [21]. To improve SVM classification accuracy, a Radial Basis Function (RBF) kernel can be used to identify the best hyperplane that separates the data points of each class with the largest margin between the classes.

F. Preforming Evaluation Metrics

There are several different methods of evaluating performance, the aim of which is to evaluate the performance of different tasks. In this study, three commonly used methods were selected for evaluating work performance [19].

- *Accuracy (ACC):* this is used to evaluate the performance of classification results for both melanoma and non-melanoma categories. Eq. (3) can be used for calculating the Accuracy:

$$\text{Acc} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

- *Sensitivity (SEN):* it is used to measure the model's ability to correctly diagnose melanoma. Eq. (4) shows how to calculate the Sensitivity:

$$\text{SEN} = \frac{TP}{TP+FN} \quad (4)$$

- *Specificity (SEP):* which is used to measure the model's ability to correctly diagnose non-melanoma. Eq. (5) that can be used for calculating SEP is:

$$\text{SEP} = \frac{TN}{TN+FP} \quad (5)$$

where:
TP refers to correctly diagnosed melanoma cases (True Positive),
FP is the diagnosed cases with melanoma incorrectly (False Positive),
TN is the diagnosed cases with non-melanoma correctly (True Negative),
FN is the diagnosed cases with non-melanoma incorrectly (False Negative).

IV. Experimental Work and Results

Step 1: The database used in this study consisted of 2357 high-resolution digital images of the Dermatoscope for malignant and benign tumors with a (.jpg) extension. These images have been collected by the International Skin Imaging Collaboration (ISIC) from the Kaggle website [22]. 2038 images are randomly selected and separated into 1433 training images and 605 test images; the train-to-test data ratio is chosen to be 70% to 30%. The training database contains 343 images diagnosed with melanoma and 1090 images diagnosed with non-melanoma. The test database contained 110 images diagnosed with melanoma and 495 images diagnosed with non-melanoma. The images in the database are of different sizes and aspect ratios, ranging from 600×450 to 6708×4439. All images in the database have been resized to 1200×900. The aspect ratio for this size is 4/3. All database images have been unified through the following two sub-steps:

- *Unsharp masking:* Using unsharp masking (3×3) to detect the borders of the lesion and to distinguish it from the skin.
- *Image resizing:* Using the bicubic interpolation technique to resize the image while preserving the fine details of the lesion and to reduce the loss of important information in the images by resizing them. Saving the resized images in a new file with the same extension format (.jpg).

Step 2: The training database is imported and converted into grayscale images for easy handling. Using the histogram equalization technique to balance the contrast in the images, which contain slight differences in texture, color, and shape that are difficult to distinguish with the naked eye.

Step 3: The process of isolating the lesion in the image was done by using structure elements and morphological operations. These processes helped to improve and refine the borders of the lesion in the image. The thresholding technique was used to segment the foreground (the skin lesion) from the background in the image. Fig. 2. shows the result of using these processes [5].

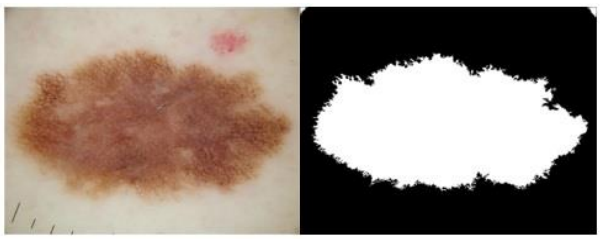


Figure 2. Results of using morphological operations and thresholding process to segment skin lesion.

Step 4: Feature extraction is done by combining the GLCM, LBP, skewness, kurtosis, and equivalent diameter to capture important information about texture, shape, and the distribution of pixel values in the image. In the GLCM, the displacement was determined by determining the distance between the two pairs of pixels, which is 2. Also, the displacement direction was determined by specifying the angle of rotation required to achieve the desired displacement, which is (00, 450, 900, 1350), corresponding to the horizontal, diagonal, vertical, and anti-diagonal directions, respectively. In the LBP, the number of neighboring pixels is (8) and the radius is (1). The high values extracted from the LBP may indicate the presence of a confused texture in the image of the skin lesion, which could be indicative of melanoma because cancerous tumors often tend to have random, irregular cells or textures. In the next step, the processed training data is saved, and characteristics are extracted from it.

Step 5: In the training phase, important features are extracted from the image dataset, capturing relevant information for melanoma detection. These features are then used to train the SVM model using the RBF kernel, which enables effective mapping of the features into a higher-dimensional space. In the testing phase, the features extracted from the test images are input into the trained model. The SVM model, utilizing the RBF kernel, transforms the test features into the same higher-dimensional space as in training. By measuring the distance of the new data points from the hyperplane, the trained model predicts the class of the test image based on the shortest distance, assigning it to either the melanoma or non-melanoma category.

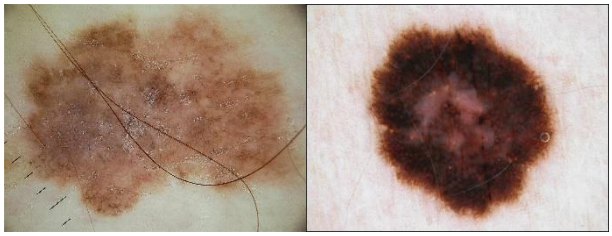
As shown in Table I, 110 images were diagnosed as Melanoma (M), and 495 images were diagnosed as Non-Melanoma (Non-M). The correct diagnosis for melanoma is TP, while FN is an incorrect diagnosis. The correct diagnosis for non-melanoma is TN, while FP is an incorrect diagnosis. The specificity of the results was 96.21%, the sensitivity was 88.34%, and the accuracy of the results in general was 94.87%. This accuracy indicates the success of training the model and its ability to distinguish between melanoma and non-melanoma.

TABLE I. The Result of Testing Dataset Images

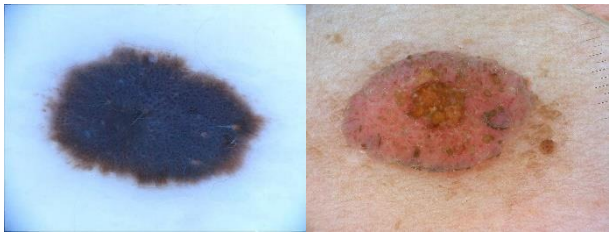
Testing (605) images	M.	Non-M.	SEP. %	SEN. %	ACC. %
M. (110)	TP (91)	TN (483)	96.21	88.34	94.87
Non-M. (495)	FN (12)	FP (19)			

Fig. 3 shows the result of True Positive (TP, 3a), False Negative (FN, 3b), True Negative (TN, 3c), and False Positive (FP, 3d).

The accuracy of the proposed system has been compared with other recent studies that used the same dataset, and the result of the comparison can be seen in Table II.



(a) (b)



(c) (d)

Figure 3. The results of applying the test images for the four types of diagnosis. (a) Image No. 103 from testing data; (b) Image No. 4 from testing data; (c) Image No. 306 from testing data; (d) Image No. 307 from testing data.

TABLE II. A COMPARISON BETWEEN THE PROPOSED METHOD AND OTHER METHODS

Methods	Classification techniques	Accuracy
Lee [23]	Logistic regression	84%
Paola [24]	CNN-Mobile Net V1	91.06%
Chin [25]	CNN-ARIMA	92.25%
Proposed method	SVM-RBF kernel	94.87%

The first group, presented in Table II [23], proposed a system, Inception V3, as transfer learning, and then the classification was tested by SVM, KNN, and logistic regression. The study showed that this method gave best accuracy of results when using logistic regression, which achieved an accuracy of 84%. The second group in Table II [24], suggested a method to train two CNN models, namely MobileNet V1 and Inception V3, and to test the accuracy of the results, MobileNet V1 and Inception V3 achieved an accuracy of 90.34% and 83.71%, respectively. While the third group in Table II [25], proposed a hybrid technical method, which was CNN-ARIMA, where the researchers explained that CNN was used in extracting features and the ARIMA model as a data classifier. This achieved an accuracy of 92.25%. The accuracy of the results achieved by the system in this study when compared with the results, which represent different methods in Table II. It can be seen that the proposed system showed success and competitive performance compared to the methods mentioned in this table, which used the same database of this study.

V. CONCLUSIONS AND FUTURE WORKS

In this study, a melanoma and non-melanoma diagnosis database was utilized to test the proposed system for the diagnosis of melanoma skin cancer. The system uses image processing techniques to increase the contrast between the lesion and the skin. Also, it uses morphological operations to remove small objects that are considered as noise in the image, such as hair or blood vessels surrounding the lesion. Combined LBP, GLCM, skewness, Kurtosis, and diameter features. These features contributed to the successful training of the SVM model. The system achieved an accuracy of 94.87% in diagnosing the test database. This result confirms the success of the system to predict the test data and prevent the spread of melanoma to the rest of the human body.

This study could be used in the future to diagnose other types of skin cancer, such as basal cell carcinoma and squamous cell carcinoma. Also using a microcontroller and programming it, this will give a portable device unit that has the ability to detect skin cancer and diagnose its type as a stand-alone platform.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Author Radwan wrote this paper with the assistance, guidance, and supervision of Nasseer; Nasseer chooses the idea of the work and contributes to assigning the scientific and theoretical background to the basic steps in this paper. All authors had approved the final version.

REFERENCES

[1] M. Ralli *et al.*, "Immunotherapy in the treatment of metastatic melanoma: Current knowledge and future directions," *J. Immunol. Res.*, vol. 2020, 9235638, 2020. doi: 10.1155/2020/9235638

[2] S. Jain, V. Jagtap, and N. Pise, "Computer aided melanoma skin cancer detection using image processing," *Procedia Comput. Sci.*, vol. 48, no. C, pp. 735–740, 2015. doi: 10.1016/j.procs.2015.04.209

[3] C. Rosendahl and A. Marozava, "Dermatoscopy and skin cancer: A handbook for hunters of skin cancer and melanoma," *Scion Publ. Ltd*, vol. 2, no. 5, p. 255, 2019.

[4] A. Kowalczyk, "Support vector machines succinctly," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2017.

[5] E. Nasr-Esfahani *et al.*, "Melanoma detection by analysis of clinical images using convolutional neural network," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, 2016, pp. 1373–1376. doi: 10.1109/EMBC.2016.7590963

[6] M. Hasan, S. D. Barman, S. Islam, and A. W. Reza, "Skin cancer detection using convolutional neural network," in *Proc. ACM Int. Conf. Proceeding Ser.*, 2019, pp. 254–258. doi: 10.1145/3330482.3330525

[7] A. Murugan, S. A. H. Nair, and K. P. S. Kumar, "Detection of skin cancer using SVM, random forest and kNN classifiers," *J. Med. Syst.*, vol. 43, no. 8, 2019. doi: 10.1007/s10916-019-1400-8

[8] M. A. A. Thaajwer and U. A. P. Ishanka, "Melanoma skin cancer detection using image processing and machine learning techniques," in *Proc. 2nd Int. Conf. Adv. Comput. Proc., ICAC 2020*, 2020, pp. 363–368. doi: 10.1109/ICAC51239.2020.9357309

[9] G. N. K. Babu and V. J. Peter, "Skin cancer detection using support vector machine with histogram of oriented gradients features," *ICTACT J. Soft Comput.*, vol. 6956, no. January, pp. 2301–2305, 2021. doi: 10.21917/ijsc.2021.0329

[10] A. M. M. Alsarraf and O. N. Ucan, "Detection and classification of melanoma by analysis of skin lesions using teledermatology," in *Proc. 2022 8th International Conference on Contemporary Information Technology and Mathematics (ICCITM)*, 2022. pp. 341–346.

[11] R. Sharma and V. Mehan, "Skin disease detection using image processing and soft computing," *ECS Transactions*, vol. 107, no. 1. pp. 17051–17061, 2022. doi: 10.1149/10701.17051ecst

[12] K. Dimililer and B. Sekeroglu, "Skin lesion classification using cnn-based transfer learning model," *Gazi University Journal of Science*, 2022. doi: 10.35378/gujs.1063289

[13] R. C. Gonzales and P. Woods, *Digital Image Processing*, 4th Ed. Pearson Education Limited, 2018.

[14] S. S. Bagade, "Use of histogram equalization in image processing for image enhancement," *Int. J. Softw. Eng. Res. Pract.*, vol. 1, no. 2, pp. 6–10, 2011.

[15] D. Chudasama, T. Patel, S. Joshi, and G. I. Prajapati, "Image segmentation using morphological operations," *Int. J. Comput. Appl.*, vol. 117, no. 18, 2015.

[16] C. Solomon and T. Breckon, *Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab*, John Wiley & Sons, 2011.

[17] A. Gautam and B. Raman, "Skin cancer classification from dermoscopic images using feature extraction methods," in *Proc. IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON*, 2020, pp. 958–963. doi: 10.1109/TENCON50793.2020.9293863
[18] M. Ramachandro, T. Daniya, and B. Saritha, "Skin cancer detection using machine learning algorithms," in *Proc. 3rd IEEE Int. Virtual Conf. Innov. Power Adv. Comput. Technol. i-PACT 2021*, 2021, pp. 5–9. doi: 10.1109/i-PACT52855.2021.9696874
[19] V. Ruthra and P. Sumathy, "Color and texture based feature extraction for classifying skin cancer using support vector machine and convolutional neural network," *Int. Res. J. Eng. Technol.*, vol. 6, issue 9, pp. 502–507, Sep. 2019.
[20] E. M. Senan and M. E. Jadhav, "Analysis of dermoscopy images by using ABCD rule for early detection of skin cancer," *Glob. Transitions Proc.*, vol. 2, no. 1, pp. 1–7, 2021. doi: 10.1016/j.gltp.2021.01.001
[21] D. Keerthana, V. Venugopal, M. K. Nath, and M. Mishra, "Hybrid convolutional neural networks with SVM classifier for classification of skin cancer," *Biomed. Eng. Adv.*, vol. 5, no. November 2022, 100069, 2023. doi: 10.1016/j.bea.2022.100069
[22] Skin Cancer ISIC | Kaggle. [Online]. Available: https://www.kaggle.com/datasets/nodoubttome/skin-cancer9-classesisic
[23] J. Z. Lee and A. P. P. Abdul Majeed, "Classification of skin cancer by means of transfer learning models," *Mekatronika*, vol. 3, no. 2. pp. 77–81, 2021. doi: 10.15282/mekatronika.v3i2.7393
[24] D. Paola *et al*. (2021). Diseño de una herramienta para la clasificación de imágenes de cáncer de piel utilizando Redes Neuronales Profundas (DNN). [Online]. pp. 65–80. Available: https://dialnet.unirioja.es/servlet/articulo?codigo=8148855
[25] C. K. Chin, D. A. B. A. Mat, and A. Y. Saleh, "Hybrid of convolutional neural network algorithm and autoregressive integrated moving average model for skin cancer classification among Malaysian," *IAES International Journal of Artificial Intelligence*, vol. 10, no. 3. pp. 707–716, 2021. doi: 10.11591/ijai.v10.i3.pp707-716