Grenze International Journal of Engineering and Technology, January Issue

Comparative Analysis of Machine Learning Classifiers for Melanoma Skin Cancer Detection

Apurva S. Solanke[1] and Prapti D. Deshmukh[2]
[1-2]Dr. G. Y. Pathrikar College of CS & IT, MGM University, Chhatrapati Sambhajinagar, India
Email: asolankeresearch@gmail.com, pdeshmukh@mgmu.ac.in

***Abstract*— Skin cancer is the most widespread and serious form of cancer in individuals. One kind of skin cancer that is severe is melanoma. If identified in its early stages, it is easily preventable. The biopsy procedure is the recommended way to diagnose and detect melanoma. This technique can be a highly time-consuming and unbearable procedure. This work provides a computer-aided detection approach for melanoma early detection. For the purpose to categorize melanoma skin cancer, we have compared various machine learning classifiers in the present study. Support vector machines (SVM), K-nearest Neighbors (KNN), Random Forest, and Naïve Bayes have all been employed for it. We used melanoma images to compare the results we obtained from these four classifiers.**

***Keywords*— SVM, KNN, Random Forest, Naïve Bayes.**

I. INTRODUCTION

The skin is the most significant organ in the human body. The muscles, bones, and internal organs underneath are shielded by the skin. The skin serves an essential role in protecting the body from UV rays. Melanin is an element found in skin cells that shields the skin from ultraviolet light. Melano-ma is the most life-threatening type of skin cancer in humans out of all of them. In the instance that melanoma skin cancer is noticed or diagnosed in its very early stages, it may be treatable, and if treatment is started sooner, it may even save the patient's life. Melanocytes, which are present in the body, are the main cause of melanoma. The biopsy procedure is the recommended way to diagnose skin cancer [1]. This method involves removing a portion of a human cell, which is then sent to a lab for analysis. To do so is the most difficult and painful procedure. For testing purposes, it will take a lot longer. More time must be dedicated to testing by both physicians and patients. Using the biopsy approach is risky since it could lead to the disease spreading to other places of the body [2].

The majority of the researchers who worked on this study recommended a number of detecting techniques. The four primary steps of the standard detection technique—pre-processing, segmentation, feature extraction, and classification—are used to identify melanoma, a kind of skin cancer [3]. Pre-processing stage helps to remove noise presents in lesion area like bubbles, hair particles. Segmentation process separates the lesion part from the skin in order to get the region of interest [4]. The feature extraction methodology of many computerized melanoma detection systems has been largely based on the ABCD rule due to its effectiveness and simplicity of implementation [5].

In this research, after the feature extraction we have applied some machine algorithms to classify melanoma lesion area. We have compared the result of different classification algorithms.

Grenze ID: 01.GIJET.12.1.44_2

A. Contribution

This paper compares four machine learning models — SVM, KNN, Random Forest, and Naive Bayes — to find the best one for detecting melanoma skin cancer. We used ac-curacy, sensitivity, and specificity to measure their performance. The results show that SVM works best for this task. This study helps in choosing the right model for better and early detection of melanoma.

II. PROPOSED METHODOLOGY

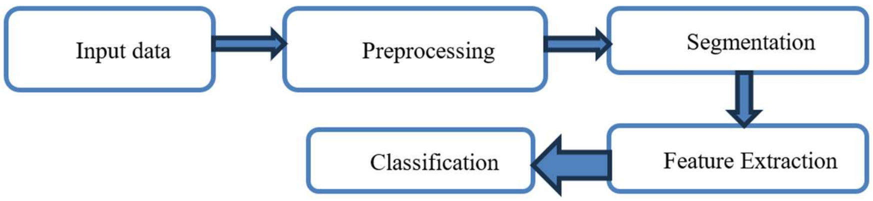
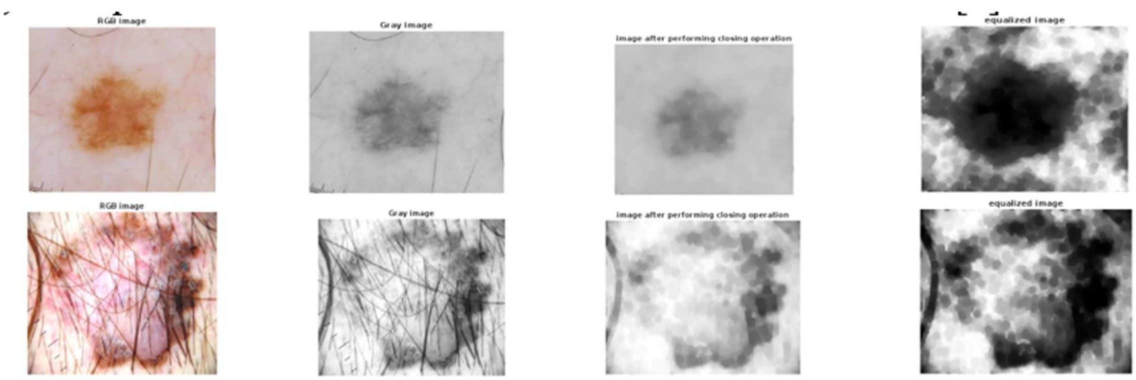


Figure 1. Flow of Methodology

A. Preprocessing

Reducing or eliminating unwanted portions of the image or background is the primary objec-tive of the preprocessing stage, which aims to improve image attributes like quality, clarity, etc. Prepro-cessing primarily consists of image improvement, noise reduction, and grayscale conversion. The first step in this suggested system is to convert all of the images to grayscale. Following morphological op-eration for hair removal, histogram equalization has been used to enhance the lesion area. [6].



Original image Grayscale Hair removed Histogram Equalization

Figure 2. Original Image and Pre-processed images

B. Segmentation

The process of isolating an image's region of interest is called segmentation. The technique of image segmentation uses the enhanced image output as its input. The lesion region has been separated from the surrounding area through segmentation. For segmentation, OTSU thresholding is utilized [7].

$$g(i,j) = \begin{cases} 1, f(i,j) \geq T \\ 0, f(i,j) \leq T \end{cases} \qquad (1)$$

Figure 3. Segmented lesion

C. Feature Extraction

The most important step in the classification process is thought to be feature extraction. Fea-ture extraction is the process of obtaining pertinent characteristics from the provided input dataset in order to carry out additional calculations like detection and classification. The ABCD rule is used to figure out the melanoma's asymmetry lesion, border irregularity, which measures the malignant margins and ranges from 0 to 8, color, which analyses the color of the skin that is visible to the naked eye, Di-ameter: The likelihood of malignancy increases if the diameter is more than 6 mm. These are the ABCD features which are considered for the identification of melanoma or non-melanoma [8] [11].
A commonly used method for the clinical diagnosis of melanoma is the ABCD rule. It stands for Asymmetry, Border irregularity, color variation and Diameter.

- Asymmetry: Measures by dividing the lesion area in half and compared the two halves. Compute the asymmetry indices for each of the two divisions.
- Border Irregularity: By calculating the perimeter and comparing it to a circular shape, one can identify irregularity.
- Color variation: Identify how many distinct colors there are in the lesion region, as well as the stand-ard deviation of color intensity.
- Diameter: Measures the diameter of the lesion.

Asymmetry Feature

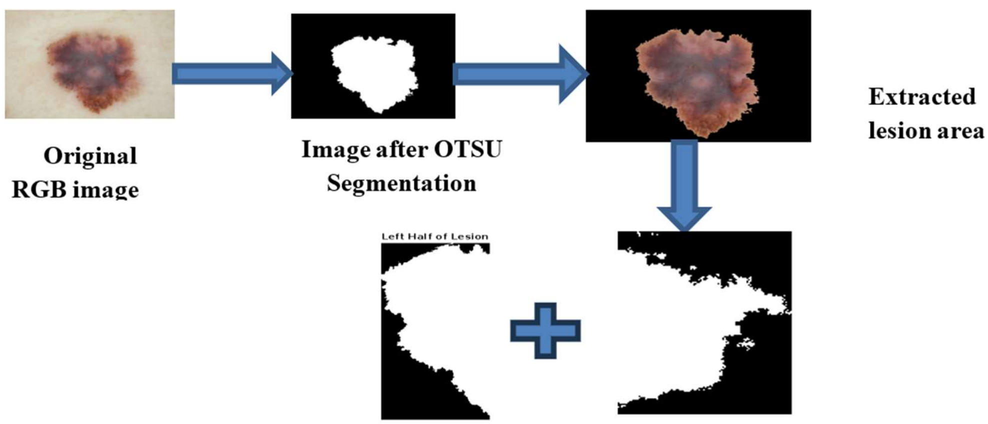
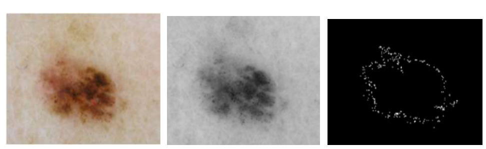


Figure 4. Asymmetry Feature

Border Feature



Original Image Gray scale Image Canny edge

Figure 5. Border feature

Circularity (border regularity): 0.41735
The lesion has an irregular (uneven) border

- Border Irregularity is a key feature of the ABCD rule used in melanoma detection.
- Irreular borders are characteristic of malignant melanomas and are usually uneven, jag-ged, or notched.

Color Feature

- Melanoma often shows varied colors, including shades of brown, black, red, blue, and even white. The colors are typically unevenly distributed within the lesion.
- Benign lesions usually have a more uniform color, typically brown or tan, without signifi-cant variation.
- Color refers to the heterogeneity of pigmentation in a lesion, with melanomas of-ten displaying multiple, uneven colors (brown, black, red, blue, white).

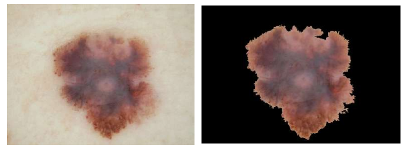


Figure 6. Color Feature

Brown: 9.78%, Red: 90.79%, Black: 0.00%, Blue: 0.00%

Diameter feature

The diameter is determined across the lesion's broadest area. A diameter larger than 6 millime-ters is usually regarded as suspicious and calls for additional research. Melanoma can also occur in tumors smaller than 6 mm, however malignancy is more frequently linked to larger lesions.

Figure 7. Diameter Feature

D. Classification

Classifying the images into benign melanoma and malignant melanoma is the last stage. The classifier is used to distinguish melanoma images from non-melanoma images. Using machine learning techniques such as SVM, KNN, Random Forest, and Naïve Bayes, der-matoscopy images are classified as either melanoma or non-melanoma.

Support Vector Machine (SVM)

The most often used classifier type is the support vector machine (SVM). SVM clas-sifiers are used on the characteristics of melanoma and benign skin lesions. A set of data is collected, and it is then predicted which two categories each input data point belongs to. A hyperplane created by the SVM separates two classes with a significant distance between them [9].

K Nearest Neighbour (KNN)

The K-Nearest Neighbor classifier is the most simple, effective, and fast. The major-ity vote of an image's neighbors determines its classification. The KNN classifier receives the training and testing samples, and uses closest distance to determine each class. Both be-nign and melanoma features are subjected to KNN classifiers [10].

Naïve Bayes

In essence, naïve Bayes analysis is predicated on the Bayes theorem-based prior probability belief. Large data is not necessary, and the qualities are fast and conditionally independent, meaning they are not dependent on one another. This is the primary benefit.

Random Forest

Random Forest is a well-liked and powerful machine learning technique for applica-tions involving regression and classification. Because it creates numerous decisions trees and aggregates their output to provide a final forecast, it is referred to as a "forest." Based on feature values, Random Forest constructs a large number of trees that perform generalized splits. These splits may not always capture the fine-grained separation between melanoma and benign cases. Random Forest may suffer in performance unless features are carefully se-lected or reduced.

III. Experimental Result

In this study, multiple machine learning algorithms were applied to classify melanoma skin cancer from dermoscopic images. The models tested include Support Vector Machine (SVM), Na-ive Bayes, K-Nearest

Neighbor (KNN), and Random Forest. Each classifier was evaluated using the same dataset to ensure a fair comparison of their performance. SVM worked well because it can clearly separate cancerous and non-cancerous lesions, even when the data is complex. Naive Bayes didn't perform as well since it assumes features are independent, which is not true for im-ages. KNN had average results and was affected by noisy data. Random Forest was good but still not as accurate as SVM. These results show that SVM is a strong choice for helping doctors detect melanoma early and accurately.

TABLE I. CLASSIFICATION OF MELANOMA WITH DIFFERENT ALGORITHMS

Image No.	**Actual Label**	**SVM Result**	**KNN Result**	**Random Forest Result**	**Naive Bayes Result**
1	Melanoma	☑ Melanoma	Melanoma	Melanoma	✗ Benign
2	Benign	☑ Benign	Benign	Benign	☑ Benign
3	Melanoma	☑ Melanoma	✗ Benign	☑ Melanoma	☑ Melanoma
4	Benign	☑ Benign	Benign	Benign	✗ Melanoma
5	Melanoma	☑ Melanoma	Melanoma	Melanoma	✗ Benign
6	Benign	☑ Benign	Benign	Benign	☑ Benign
7	Melanoma	☑ Melanoma	Melanoma	Melanoma	☑ Melanoma
8	Benign	☑ Benign	✗ Melanoma	☑ Benign	☑ Benign
9	Melanoma	☑ Melanoma	✗ Benign	☑ Melanoma	☑ Melanoma
10	Benign	☑ Benign	Benign	Benign	✗ Melanoma
11	Melanoma	☑ Melanoma	Melanoma	Melanoma	✗ Benign
12	Benign	☑ Benign	Benign	Benign	☑ Benign
13	Melanoma	☑ Melanoma	✗ Benign	☑ Melanoma	☑ Melanoma
14	Benign	☑ Benign	Benign	Benign	✗ Melanoma
15	Melanoma	☑ Melanoma	Melanoma	Melanoma	☑ Melanoma
16	Benign	☑ Benign	Benign	Benign	☑ Benign
17	Melanoma	☑ Melanoma	✗ Benign	☑ Melanoma	✗ Benign
18	Benign	✗ Melanoma	Benign	Benign	✗ Melanoma
19	Melanoma	☑ Melanoma	Melanoma	Melanoma	☑ Melanoma
20	Benign	☑ Benign	Benign	Benign	☑ Benign

TABLE II. PERFORMANCE MEASURES

Classification techniques	Sensitivity (%)	Specificity (%)	Accuracy (%)
Random Forest	89.5	91.2	90.3
SVM	93.8	94.1	98
K-Nearest Neighbor	87.4	89	88.2
Naive Bayes	82.1	85.7	83.6

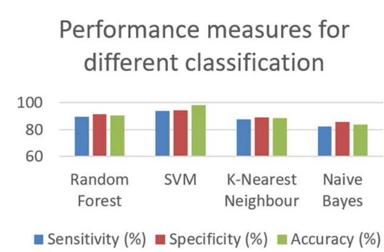


Figure 8. Graphical representation of performance measures

TABLE III: ACCURACY OF DIFFERENT CLASSIFICATION TECHNIQUES

Image No.	Random Forest Accuracy (%)	SVM Accuracy (%)	KNN Accuracy (%)	Naive Bayes Accuracy (%)
1	90	**92**	88	85
2	91	**93**	89	86
3	89	**91**	87	84
4	92	**94**	90	86.5
5	88	**90**	86	83
6	93	**95**	91	87
7	90.5	**92.5**	88.5	85.5
8	91.5	**93.5**	89.5	86.5
9	89.5	**91.5**	87.5	84.5
10	92.5	**94.5**	90.5	87.5

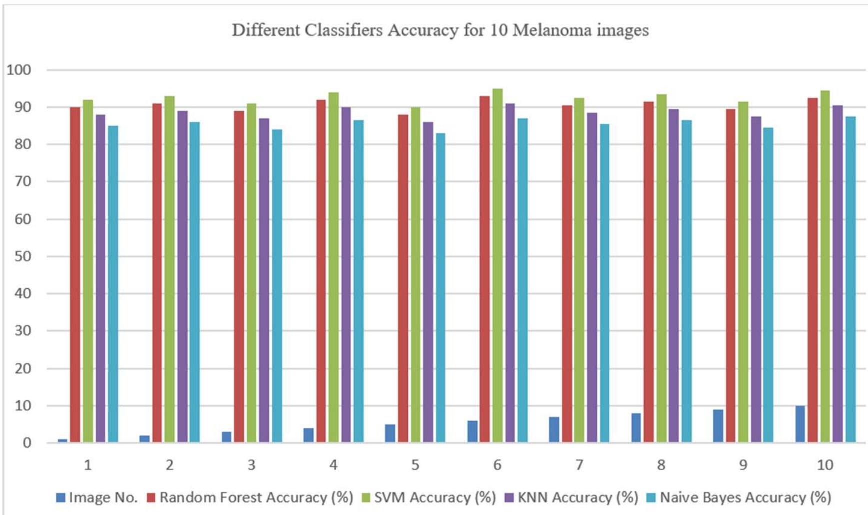


Figure 9. Graphical representation of different classification techniques

IV. CONCLUSION

This study presented a comparative analysis of four prominent machine learning classifiers-SVM, KNN, Random Forest and Naïve Bayes based on the performance in classification tasks. The models were evaluated using key performance metrics: accuracy, sensitivity and specificity. The results clearly indicate that support vector machine (SVM) outperformed the other classifiers, delivering the highest accuracy and sensitivity, making it highly effective for correctly identifying true positive cases. The primary reason is that SVM is highly effective in handling high-dimensional and complex feature spaces, which are common in medical image analysis. It excels at finding an optimal decision boundary between malignant and benign classes, even with smaller datasets — a typical scenario in medical research. Unlike Naive Bayes, which assumes feature independence (often untrue in image data), SVM can capture relationships between correlated features such as texture, asymmetry, and colour. KNN, although simple and intuitive, is sensitive to noise and performs poorly with high-dimensional data, while Random Forest can sometimes overfit and requires careful tuning of multiple parameters. SVM, on the other hand, offers a balanced and robust solution with fewer hyperparameters and consistently high accuracy. Therefore, it is concluded that SVM is the most reliable and accurate classifier for melanoma detection among the models tested in this study.

REFERENCES

[1] Thaajwer, M. A., & Ishanka, U. P. (2020). Melanoma skin cancer detection using image pro-cessing and machine learning techniques. IEEE Xplore, 363–368. https://doi.org/10.1109/icac51239.2020.9357309.

[2] B. A. Uzma and T. Sarode, "Skin Cancer Detection Using Image Processing," International Re-search Journal of Engineering and Technology (IRJET), vol. 04, no. 04, pp. 2875-2881, 2017.

[3] Y. Vikash and D. Vandana, "A study on automatic early detection of skin cancer," Int. J. Ad-vanced Intelligence Paradigms, vol. 12, no. 3/4, pp. 392-399, 2019.

[4] J. Shivangi, j. Vandana and P. Nitin, "Computer aided Melanoma skin cancer detection using Image Processing," International Conference on Intelligent Computing, Communication & Convergence, pp. 735-740, 2015.
[5] G. P. Asha, Anitha.J and P. Jacinth, "IDENTIFICATION OF MELANOMA IN DERMOSCOPY IMAGES USING IMAGE PROCESSING ALGORITHMS," 2018 International Conference on Control, Power, Communication and Computing Technologies (ICCPCCT), pp. 553-557, 2018.
[6] M. Krishna Monika, N. Arun Vignesh, Ch. Usha Kumari a, M.N.V.S.S. Kumar, E. Laxmi Lydia "Skin cancer detection and classification using machine learning", 2020.
[7] Vijayalakshmi M M, "Skin cancer detection and classification using machine learning", Inter-national Journal of Trend in Scientific Research and Development (IJTSRD), Volume: 3 | Issue: 4 | May-Jun 2019.
[8] Senan, E. M. & Jadhav, M. E. Analysis of dermoscopy images by using ABCD rule f or early detection of skin cancer. Glob. Transit. Proc. 2(1), 1–7 (2021).
[9] Arora, G., Dubey, A. K., Jafery, Z. A. & Rocha, A. Bag of feature and support vector machine based early diagnosis of skin cancer. Neural Compute. Appl. 34, 8385–8392 (2022)
[10] P. Kavitha Research Scholar, Dr. V. Jayalakshmi Professor "Survey of Skin Cancer Detection using Various Image Processing Techniques" Third International Conference on Intelligent Sustainable Systems [ICISS 2020].
[11] Vishala B M , Kavitha Sooda "Melanoma Skin Cancer Detection with Machine Learning" In-ternational Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429 october 2020.
[12] Nadia Smaoui Zghal, and Nabil Derbel "Melanoma Skin Cancer Detection based on Image Processing" Current Medical Imaging, 2020, Vol. 16, No. 1.
[13] A Pushpalatha , Dharani P, Dharini R , Gowsalya J "Skin Cancer Classification Detection using CNN and SVM" ICCCEBS 2021.
[14] Praveen Banasode, Minal Patil, Nikhil Ammanagi "A Melanoma Skin Cancer Detection Using Machine Learning Technique: Support Vector Machine" ICFEST 2020.