IJICIS, Vol.24, No.1, 12-19
DOI: 10.21608/ijicis.2024.260141.1313

International Journal of Intelligent Computing and Information Sciences

https://ijicis.journals.ekb.eg/

MELANOMA OF THE SKIN CANCER DIAGNOSIS USING SUPPORT VECTOR MACHINE

Zeina Rayan*

Computer Science Department,
Faculty of Computer and Information Sciences, Ain Shams University,
Cairo, Egypt
zeinarayan@cis.asu.edu.eg

Islam Hegazy

Computer Science Department,
Faculty of Computer and Information Sciences, Ain Shams University,
Cairo, Egypt
islheg@cis.asu.edu.eg

Mohamed Roushdy

Computer Science Department,
Faculty of Computer and Information Sciences, Ain Shams University,
Cairo, Egypt
mroushdy@cis.asu.edu.eg

Abdel-Badeeh M. Salem

Computer Science Department,
Faculty of Computer and Information Sciences, Ain Shams University,
Cairo, Egypt
absalem@cis.asu.edu.eg

Received 2024-01-02; Revised 2024-03-02; Accepted 2024-03-70

Abstract: *Cancer is one of the diseases, caused by cell divisions that can be fatal. It has been the second most common cause of death for the past period of years globally. Any region of the body can be affected by a wide range of disorders collectively referred to as cancer. Neoplasms as well as malignant tumors are other terms that are used for describing cancer. Skin is one of the body parts that can be affected. Early melanoma skin cancer detection is a must so that the mortality rate of skin cancer patients is decreased. The accuracy of early detection of melanoma skin cancer can be enhanced through applying machine learning methods. This paper provides a model that can detect melanoma skin cancer early. This model is built using the dataset that the International Skin Imaging Collaboration has provided. The proposed model using support vector machines achieved a promising accuracy of 95.96%.*

Keywords: *Artificial Intelligence, Smart Health, Medical Informatics, Support Vector Machine, Convolution Neural Network.*

1. Introduction

Cancer is one of the top leading causes for death around the globe, responsible for nearly ten million deaths in the year 2020 [1]. Skin Cancer is one of the top leading causes of death related to the Cancer diseases. The most common new cases of skin cancer in 2020 was 1.20 million cases [2]. Melanoma

***Corresponding Author**: Zeina Rayan

Computer Science Department, Faculty of Computer and Information Science, Ain Shams University, Cairo, Egypt

Email address: zeinarayan@cis.asu.edu.eg

skin cancer is considered one of the dangerous types of cancer that may lead to death, and it has a high ability to attack other organs, and tissues [3].

Machine Learning (ML) is one of the areas under Artificial Intelligence (AI) [4]. Creating and refining algorithms that allow computer systems to adapt their behavior based on actual data is the focus of machine learning [5]. AI, and ML can help in early detection and prediction of skin cancer using skin imageries effectively [6] [7], such a system is also called Computer-Aided Diagnostic systems (CAD).

Melanoma skin cancer can be treated carefully, if detected early. However, it is possible to improve pathologists' performance and increase the treatment success rate by using Machine learning. The low contrast and high degree of similarity between areas of normal skin and areas of skin lesions present considerable obstacles for automatically detecting melanoma skin cancer with the power of machine learning techniques [8]. So, at the moment, research into automatic skin cancer detection is regarded enthusiastic.

The structure of this paper is as follows: section 2 provides a comparison analysis of various machine learning techniques that are applied for skin cancer detection, section 3 contains the proposed methodology, section 4 shows the discussion and outcomes. Finally, section 5 concludes by outlining the conclusions and next steps.

2. Related Work

This section constructs a study comparing several ML techniques which are used in order to detect melanoma skin cancer.

Ramya et al. [9], suggested a model based on Support Vector Machine (SVM), and the “adaptive histogram equalization”, and “wiener filter” for pre-processing. They used “active contour segmentation” for segmentation. The attributes used for the model were utilized using “Gray Level Co-occurrence Matrix (GLCM)”. They used the “International Skin Imaging Collaboration (ISIC)” dataset. They obtained an accuracy of 95%.

Premaladha et al. [10], proposed a smart ensemble model for categorization and prediction of skin cancer based on “neural networks” using Deep Learning (DL) and “hybrid AdaBoost SVM”. They used “Median filter” and “Contrast Limited Adaptive Histogram Equalization” for pre-processing and image enhancement. They used “Normalized Otsu’s Segmentation” algorithm for segmentation. Their model used the Skin Cancer (SC) and Benign Tumor Image Atlas dataset. The model reached 91% accuracy.

Bareiro Paniagua et al. [11] proposed a methodology based on SVM. They used the PH^2 “a public database for the analysis of dermoscopic images” dataset. The attributes were obtained using “ABCD rule”, for the segmentation phase, they used “lesion segmentation” technique. Their model achieved 89% accuracy.

Aima et al. [12], proposed a smart approach based on “Convolution Neural Network (CNN)”. Image processing is applied for feature extraction, before this image is passed to CNN. Their approach obtained an accuracy of 73% on the ISIC dataset.

Dai et al. [13], proposed a model based on CNN. Local work was done on the location where the test data was stored to implement the classification process. Their work was done on a huge gathering of multi-source datasets of skin imageries. Their model obtained reduced latency, less power, and confidentiality was improved along with 74% accuracy.

Zghal, and Derbel [14], proposed a model based on the total of the four retrieved parameters times their respective weights yielding the “Total Dermoscopy Value (TDV)”. They used filtering and contrast enhancing algorithms for pre-processing, for segmentation they used lesion segmentation. They used the PH^2 dataset, and their model achieved an accuracy of 90%.

Lingaraj, et. al. [15], proposed a model based on SVM namely “Veritable Support Vector Machine (VSVM)”. Their model focuses on extracting features from the medical imagery of the datasets. Their model achieved different accuracies when applied to different datasets, where it obtained an accuracy of 83.74% on the “Health Information System (HIS)” dataset, and 89.05% on the ISIC dataset.

Diame, et. al. [3], proposed a model based on an ensemble of Resnet50 which is “a convolutional neural network that is 50 layers deep” [16], and U-Net which is “convolutional neural network that was developed for biomedical image segmentation” [17]; which can be called as Resnet50_unet. As for pre-processing they used lesion segmentation. Their model was built and tested using ISIC dataset and it achieved an accuracy of 95.76%.

Table 1 shows a study of comparison of several machine learning methods used for the diagnosis of melanoma skin cancer.

Table 1 Machine Learning Methods for Skin Cancer Diagnosis

Authors	**Year**	**Dataset**	**Segmentation**	**Machine Learning Technique**	**Accuracy**
Ramya, et al	2015	ISIC	Active Contour Segmentation	GLCM and SVM	94%
Premaladha, et al.	2016	SC and Benign Tumor Image Atlas	Normalized Otsu’s Segmentation	ANN and ensemble AdaBoost SVM	91%
Bareiro Paniagua, et al.	2016	PH2	Lesion Segmentation	SVM	89%
Aima, et al.	2019	ISIC	-	ANN	73%
Dai, et al.	2019	Huge gathering of multi-Source skin imageries	-	CNN	74%
Zghal, and Derbel	2020	PH2	Lesion Segmentation	Image Processing	90%
Lingaraj, et. al.	2021	HIS, ISIC	-	VSVM	83.74%, 89.05%
Diame, et. al.	2022	ISIC	Lesion Segmentation	Resnet50_unet	95.76%

From table 1, we can conclude that the accuracy is always increased when segmentation algorithm is used before moving forward to the machine learning phase. Also, there are various machine learning techniques that can be used for melanoma skin cancer detection, such as GLCM, CNN, SVM …etc. However, the "support vector machine" is the most commonly used algorithm for melanoma skin cancer detection.

3. The Proposed Model

Our research aims to create a machine learning model that can classify skin cancer (Melanoma) efficiently. The developed model is applied to the International Skin Imaging Collaboration dataset. Figure 1 presents this model.

The model consists of three main steps. The first step in our model is data pre-processing. The second step is feature extraction. The third step is to apply a supervised machine learning technique, and then evaluating the model using the accuracy performance metric. Each step will be explained in detail in the next following sub-sections.

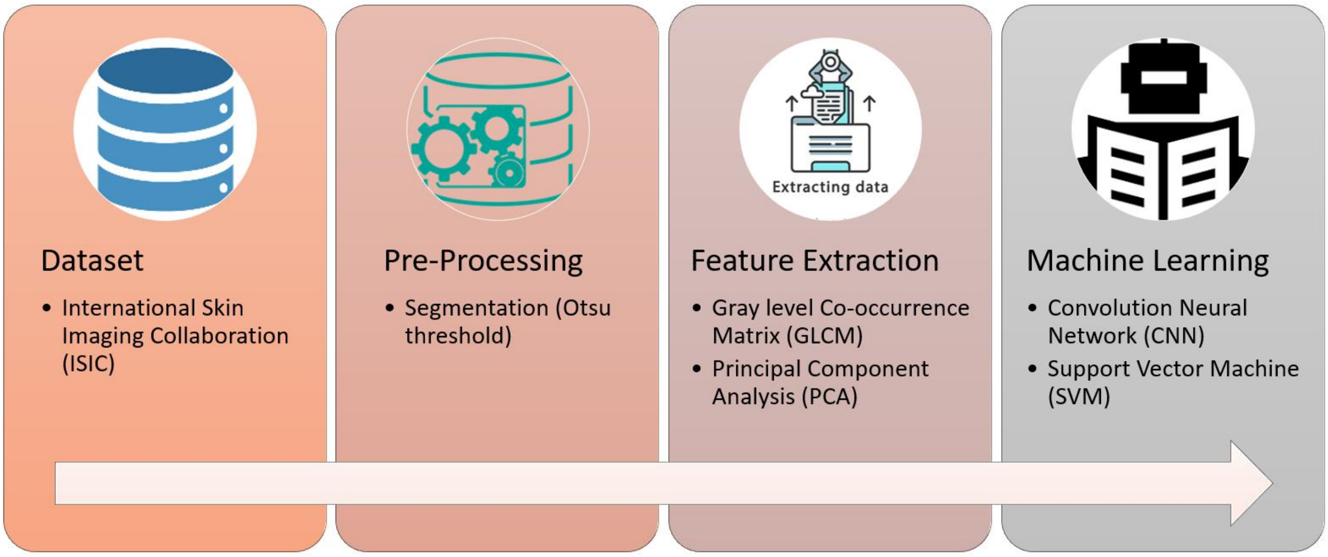

Figure 1 The Proposed Classification Methodology

3.1. Dataset

The dataset is provided by the "International Skin Imaging Collaboration (ISIC)". ISIC is a collaboration between academia and business aimed at easing the use of digital skin imaging to reduce the melanoma death rate [18].

The dataset is publicly available online. It contains a total of 2000 records. 374 records for melanoma, and 1626 records for non-melanoma. In this dataset, each image has a different resolution that needs to be taken into account. This is why a pre-processing phase had to be done. The ISIC dataset description is presented in Table 2.

Figure 2 shows some melanoma samples, and Figure 3 shows some of non-melanoma samples from the used dataset.

Table 2 ISIC dataset information

Classification Term	No. of Images	Label Number
“Melanoma”	374	1
“Non-Melanoma”	1626	2

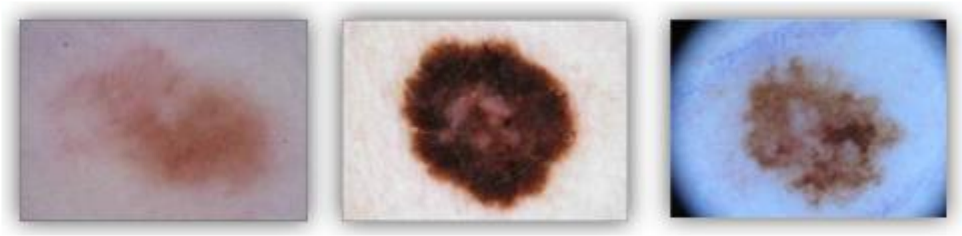

Figure 2 dataset’s melanoma samples.

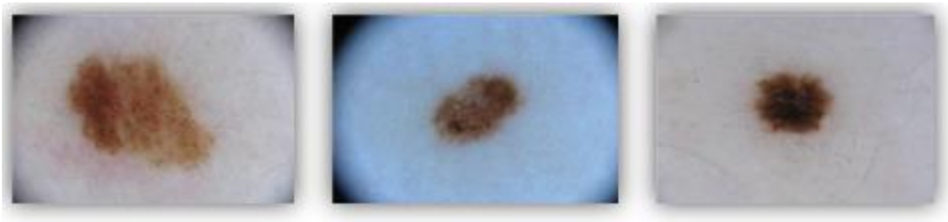

Figure 3 dataset’s non-melanoma samples.

3.2. Pre-Processing

Image segmentation is an important phase in image processing. It is responsible for dividing an image into more than one region or segment based on some criteria [19]. More specifically, one of the most widely used techniques for image segmentation is the Otsu’s technique [20]. In our work, the accuracy increased when the images are segmented using Otsu threshold algorithm. In “Otsu’s Segmentation”, threshold is obtained. The image pixels are replaced by one if its intensity value is greater than a fixed threshold and a zero replaces the rest values. Determining the optimal threshold for Otsu's method often involves experimentation and validation. In our work, we conducted multiple experiments by trying different threshold values and assessing the segmentation performance using personal validation and then also the accuracy metric.

3.3. Feature Extraction

Feature extraction is useful when dealing with images; as there is a need to reduce the number of attributes that will be used as an input for the machine learning phase [21] [22]. Two of the most popular techniques for feature extraction that are used for images are “GLCM”, and “Principal Component Analysis (PCA)”.

In this paper these two algorithms were applied but only one proved to be more efficient as it reached a higher accuracy. The PCA feature extraction algorithm is the one that reached a better accuracy.

3.4. Machine Learning

In this paper we tried two different machine learning techniques to detect melanoma skin cancer. The first machine learning technique is the Convolution Neural Network (CNN). The second machine leaning technique is “Support Vector Machine (SVM)”. SVM is a method of supervised machine learning. Its objective is to construct a hyperplane that, in an N-dimensional space with N being the number of features, can unambiguously classify the dataset samples. To convert the input features to a different representation space, SVM uses a kernel function. Commonly used kernels include "polynomial function," "linear function," and "radial basis function (RBF)". Still, anyone can also specify his own kernels [23].

CNN is a supervised deep neural network with the main objective of classifying images in several predefined classes which are melanoma, non-melanoma. It can also be used for feature extraction. So, in our work there is no pre-processing, and no feature extraction is done as CNN can be also used for feature extraction.
In case of the SVM: segmentation is done as a pre-processing using Otsu threshold. For the feature extraction we tried two methods: the first one is “Gray Level Co-occurrence Matrix (GLCM)” [24], and the second one is “Principal Component Analysis (PCA)” [25].

4. Outcomes and Discussion

In this paper two different classifiers were used to conduct experiments, the first classifier is CNN with linear activation function, and the second classifier is SVM. The dataset was divided into a training set and a testing set, where the testing subset is 20% of the original dataset and the training subset contains 80% of the original dataset.
The proposed methodology was evaluated using the accuracy performance metric. The ratio of the number of correctly predicted observations, also known as True Positives (TP), to the total number of observations is referred to as accuracy, as shown in equation (1) [26].

$$\boldsymbol{Accuracy = \frac{TP}{\#Observations}} \quad (1)$$

The accuracy metric is commonly used for measuring performance in classification tasks for four main reasons. First reason: intuitive interpretation, accuracy is easy to understand and interpret. Second reason: universal applicability, where accuracy is a universal metric that can be used across different types of classification problems regardless of class distribution or imbalance. It treats all classes equally and provides a single numerical value that summarizes the overall performance of the classifier. Third reason: direct comparison, accuracy allows for straightforward comparisons between different models or variations of the same model. Finally, metric stability, in many scenarios, accuracy remains stable even when the class distribution changes, making it robust to variations in the dataset. This stability makes accuracy a reliable metric for evaluating model performance across different datasets or applications.
Table 3 shows the results conducted by the CNN classifier.

Table 3 CNN with linear activation function Results

No. of Epochs	**Accuracy**
30	56%
20	60%
10	62%

From table 3, we get that the best experiment done using CNN is when using 10 number of epochs where it achieved an accuracy of 62%. The accuracy achieved is relatively low because the dataset can be considered as a small dataset for being used in deep learning models.
However, table 4 shows the results conducted by using SVM as a classifier using different feature extraction algorithms, and different kernel functions.

Table 4 SVM Experiments and Results

Segmentation	**Feature Extraction**	**SVM Kernel Function**	**Accuracy**
-	-	RBF	44%
-	-	linear	88%
-	-	poly	88%
Otsu threshold	GLCM	linear	84.6%
Otsu threshold	PCA	linear	50%
Otsu threshold	PCA	RBF	45%
Otsu threshold	PCA	poly	65%
Otsu threshold	PCA	linear	88%
Otsu threshold	PCA	poly	80%
Otsu threshold	PCA	RBF	95.96%
Otsu threshold	PCA	linear	93.48%
Otsu threshold	PCA	poly	68.60%

From table 4, we get that the best experiment done using the SVM is when segmenting the image using Otsu threshold, extracting features using PCA, and using radial basis function as SVM's kernel function where it achieved an accuracy of 95.96%.

5. Conclusion and Future Work

Melanoma skin cancer detection is a difficult issue that has persisted for many years of research since it frequently does not show symptoms until much later. The model proposed in this paper for melanoma skin cancer detection demonstrated a detection performance of 95.96% accuracy. The proposed model can easily be applied to improve clinical decision making. In future work, we will work on improving the accuracy using CNN, through working on multi-source datasets, and trying different activation functions. Also, we will help develop a smartphone application that will aid in early identifying "melanoma skin cancer".

References

[1] J. Ferlay, M. Ervik, F. Lam, M. Colombet, L. Mery, M. Piñeros, A. Znaor, I. Soerjomataram and F. Bray, "Global cancer observatory: cancer today. Lyon: International Agency for Research on Cancer," in *https://gco.iarc.fr/today (accessed February 2021)*, 2020.

[2] World Health Organization (WHO), "Cancer," in *https://www.who.int/news-room/fact-sheets/detail/cancer (accessed December 2021)*, 2021.

[3] Z. Diame, M. ElBery, M. Salem and M. Roushdy, "Experimental Comparative Study on Autoencoder Performance for Aided Melanoma Skin Disease Recognition," *International Journal of Intelligent Computing and Information Sciences,* vol. 22, no. 1, pp. 88-97, 2022.

[4] Z. Rayan, M. Alfonse and A.-B. M. Salem, "Machine learning approaches in smart health," *Procedia Computer Science,* vol. 154, pp. 361-368, 2019.

[5] J. D. Kelleher, B. M. Namee and A. D'arcy, Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies, MIT Press, 2020.

[6] B. C., C. ME. and M. JS., "A survey of feature extraction in dermoscopy image analysis of skin cancer," *IEEE J Biomed Health Inform,* vol. 23, no. 3, p. 1096–1109, 2018.

[7] R. A., A. N., S. T., R. SIU., M. Z. and K. K., "Classifcation of acute lymphoblastic leukaemia using deep learning. Microsc Res Tech," vol. 81, no. 11, p. 1310–1317, 2018.

[8] A. Adegun and S. Viriri, "An enhanced deep learning framework for skin lesions segmentation," in *International conference on computational collective intelligence*, Cham, 2019.

[9] R. VJ., N. J., P. R. and K. LA., "Detection of melanoma skin cancer using digital camera images," *ARPN J Eng Appl Sci,* vol. 10, p. 3082–3085, 2015.

[10] P. J. and R. K., "Novel approaches for diagnosing melanoma skin lesions through supervised and deep learning algorithms," *J Med Syst,* vol. 40, p. 96, 2016.

[11] B. P. LR., L. C. DN., P.-R. DP., V. NogueraJL, T. S. and L. A., "Computerized medical diagnosis of melanocytic lesionsbased on the ABCD approach," *Clei Electron J,* vol. 19, p. 6–6, 2016.

[12] A. A. and S. AK., "Predictive approach for melanoma skin Cancer detection using CNN," *Ssrn Electron J ,* vol. Available at SSRN 3352407, 2019.

[13] D. X., S. I., M. B., C. S. and A. F., "Machine learning on mobile:an on-device inference app for skin cancer detection," in *Fourth inter-national conference on fog and mobile edge computing (FMEC)*, 2019.

[14] N. S. Zghal and N. Derbel, "Melanoma Skin Cancer Detection based on Image Processing," *Current Medical Imaging,* vol. 16, pp. 50-58, 2020.

[15] M. Lingaraj, A. Senthilkumar and J. Ramkumar, "Prediction of Melanoma Skin Cancer Using Veritable Support Vector Machine," *Annals of R.S.C.B.,* vol. 25, no. 4, p. 2623 – 2636, 2021.

[16] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.

[17] O. Ronneberger, P. Fischer and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, Germany, 2015.

[18] ISIC, "The International Skin Imaging Collaboration," in *https://www.isic-archive.com/#!/topWithHeader/wideContentTop/main (accessed january 2022)*, 2022.

[19] N. Senthilkumaran and R. Rajesh, "Image segmentation-a survey of soft computing approaches," in *International Conference on Advances in Recent Technologies in Communication and Computing*, 2009.

[20] E. Anjna and E. Kaur, "Review of image segmentation technique," *International Journal of Advanced Research in Computer Science,* vol. 8, no. 4, pp. 36-39, 2017.

[21] I. Guyon and Elisseeff, "An introduction to feature extraction," *Feature extraction: foundations and applications,* pp. 1-25, 2006.

[22] I. Guyon, S. Gunn, M. Nikravesh and L. e. Zadeh, Feature extraction: foundations and applications, vol. 207, Springer, 2008.

[23] T.-F. Wu, C.-J. Lin and R. C. Weng, "Probability estimates for multi-class classification by pairwise coupling," *Journal of Machine Learning Research,* vol. 5, no. Aug, pp. 975-1005, 2004.

[24] M. Partio, B. Cramariuc, M. Gabbouj and A. Visa, "Rock texture retrieval using gray level co-occurrence matrix," in *Proc. of 5th Nordic Signal Processing Symposium*, 2002.

[25] S. Wold, K. Esbensen and P. Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems,* vol. 2, no. 1-3, pp. 37-52, 1987.

[26] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information processing & management,* vol. 45, no. 4, pp. 427-437, 2009.