A Melanoma Skin Cancer Diagnosis Using Hybrid Feature-Optimized MSVM Classification Model On Dermotoscopic Images

Aman Shakya[1] , Mr.Anuj Kumar[2]Dr Dharamveer Singh[3]

1. Student, RD Engineering College, Ghaziabad,
2. Assistant Professor, RD Engineering College, Ghaziabad
3. Professor, RD Engineering College, Ghaziabad

Abstract:

With the current advancements in the medical field, skin cancer is measured as a simple infection in the human body. Though the existence of melanoma disease is shown as a form of cancer, it is limitations in classifying it. If Melanoma disease and some other skin lesions are verified in the initial phase signs and symptoms, prediction can be effectively attained to treat them. This dermoscopic skin image plays a significant role in diagnosing a type of skin disease precisely and rapidly. The use of the proposed method is to enhance the SCD (skin cancer detection)SN, SP, Acc rate in dermoscopic images. The research article defines an enhanced plan to detect three skin cancer image categories in early phases. The mentioned input is anSC image which, by using the research technique, the planned system would be classified into cancer or normal categories of images. The clustering method has introduced the segmentation process to divide homogeneous image edges. The image preprocessing steps are done using different steps, such as the filter method, to improve the image attributes. At the same time, the other feature sets are assessed by implementing the RGB color model. GLCM and KPCA feature extraction methods altogether. For classification, MSVM is trained using the Hybrid-featured-optimized MSVM method. Several feature sets are precisely calculated to attain a better outcome using the skin cancer dermoscopic image database HAM10000. The novel work advises that hybrid-featured-optimize MSVM best compared with the other methods, efficiently predicts SC and creates an acc. rate of 98.0 percent. The outcomes are extremely precisely compared to other methods in a similar field.

Keywords*: *Skin Cancer Disease, MELANOMA, Image Segmentation using RGB color-space model, Hybrid-featured-optimize MSVM, DE-ANN existing model.

1. Introduction

Cancer is a main healthcare problem throughout the world. Generally, global statistics (GSs) advise 10.0 million expiries due to cancer in 2020 [1]. The commonly identified cancers include breast cancer (BC) in females, lung cancer (LC), and prostate cancers (PCs). Stomach, Liver, and Lung tumors are the main sponsors of tumor deaths [2]. SC includes both malignant melanoma (MM) and non-melanoma SC,typical cancers in CAUCASIANS, and their occurrence is growing. According to the United States SC foundation, SC affects more people in the US each year than all other lesions combined [3]. Melanoma is the worst stage of SC. If analyzed initially, it may be treated successfully with surgical processes. But, once there is meta-stasis, existence rates are optimized expressively [4]—analysis of melanoma disease based on the medical test and searchings on the cancer biopsy.

Early-stage detection is a key to the reliable treatment and improved results of SCs. Doctors can precisely analyzethe lesion by seeing their limited numbers. There is a requirement to design Ass (automated systems) toexplore the disease effectively to protect lives and optimize financial and health burdens on the patients. SCs has not easy to identify from joint cancer skin diseases, and melanoma has a specific diverseaspect. ML (machine learning) may aid in the ESD (early-stage detection) of SC, minimizing the loadmortality connected with the infection. In addition, to optimizing the burden, the ML-based system has helped by enhancing the SLD (skin lesion diagnostics). ML is an AI (artificial intelligence) method connecting SMs (statistical methods) and techniques that can increasingly study information to classify the features of novel samples and perform the desired work. So, the problematicways are calculated to perform the tasks that are otherwise not easy to understand in the human brain [5].

CAD (computer-aided-diagnostics) [6] developments upcomingguidelines to study medical images (MIs) using DIP (digital-image-processing). These methods supportfetching the reliable feature sets from the MIs, including color matrices, shape, texture, asymmetry, etc. The detection of cancer-based on these feature sets using ML-based techniques. In computer-based techniques, the main phases are the SL (skin lesion) image acquisition, segmentation of SC image regions, feature extraction and selection, classification, etc. There happen several classification techniques that are utilized to detect SL.

The existing work has implemented differential evolution with an artificial neural network (ANN) to generate enhanced and detect the skin cancer images. The outcome defines that the research method is advanced toward automatic segmentation and feature extraction using the HAM10000 dataset [7]; the accuracy rate of the existing process is near around 97.4 percent. The proposed work defines a research method to verify image feature sets and improve feature selection and classification precision. The proposed work implemented grey wolf optimization (GWO) with a multi-class support vector machine (MSVM) to produce enhanced grouping on DIs (dermoscopic images). Texture featuresets are extracted using GLCM and KPCA, while color features

are removed using the RGB method. Typically, the Hybrid Feature-Optimized MSVM classification model is executed to identify image edges based on reliable feature sets. The research technique effectively classifies cancer and non-cancer cells with maximum precision. The outcome defines that the research method is advanced toward image segmentation and feature extraction using the HAM10000 and PH^2 [8] dataset; the accuracy rate of the proposed process is near around 98.0 percent.
The research paper is described in five sections. Section 2 is the related work defined by the proposed work in section 3. Section 4 defines the output of the proposed method, and section 5 accomplishes the proposed work.

2. Literature Review

This section discussed skin cancer disease detection using ML and DL methods. Surveyed several articles are concluded that SVM and CNN classification methods mostly preferred to detect the disease in skin cancer. **Nawar, A., Sabuz, et al. (2021) [9]** developed a model for quickly and inexpensively identifying issues related to skin disease. The proposed technique was mainly an image processing method. The suggested system could identify different forms of skin illnesses based on the feature extraction that used chromatic segmentation methods and an SVM classification algorithm. With a prediction performance of 94.79 percent, the proposed approach effectively recognizes eight different skin conditions. The suggested concept was simple, quick, and works with any configurable platform, including desktops, Android smartphones, tablets, etc.**Elngar, A. A., Kumar, et al. (2021) [10]** presented a system for developing an android mobile application that integrates SVM with CNN classifiers. As a result, numerous experiments on the database were carried out to analyze the effectiveness of the suggested system. This database included over 3000 files gathered from many sources such as Cairo university hospital, Beni-Suef university hospital, and numerous web pages to become more realistic and accurate. A comparison of various feature extraction techniques with various classifiers was carried out.**Srinivasu, P. N., et al., (2021) [11]** presented a DL-dependent MobileNet V2 and LSTM-based automated process for detecting skin conditions. The MobileNet V2 framework is considered more efficient and accurate, and it could be used on minimal computational units. For exact estimations, the suggested framework helped maintain domain-specific data. The progression of pathological development was measured using a GLCM (Grey-level co-occurrence matrix). The results were compared to those of other cutting-edge frameworks like CNN, FTNN, VGG, and DCNN for LSIR (Large-Scale Image Recognition) design that extended with minimal alterations. The suggested technique surpassed traditional approaches with over 85 percent accuracy using the HAM10000 database.Using convolutional neural networks, **Naeem, A., et al., (2020) [12]** offered a structured survey on melanoma skin cancer classification. Only convolutional neural network categorizers were explored in detail, and the validity of such categorizers was compared when tested on unreleased databases. The main goal of such a survey was to compile a modern survey that could recognize the newest trends in surveys, issues, and alternatives for melanoma detection and analyze current melanoma cancer identification solutions using DL. In addition, the presented melanoma identification taxonomy was examined, which highlighted the main variances of current melanoma identification methods. Finally, recommended techniques, difficulties, and solutions were presented, beneficial to researchers working in melanoma detection.**Kumar, M., et al., (2020) [13]** presented computer-based approaches for correctly identifying early indications of three types of skin cancer. This survey categorized SC by employing DE-ANN; this survey categorized SC (Skin Cancer). Several filters were utilized to improve the features of a picture in pre-processing. In contrast, other attributes were evaluated by executing RGB color-space, LBP, and GLCM techniques that greatly contributed to the skin ailment categorization. Different attributes were correctly evaluated for better outcomes by utilizing skin cancer picture databases HAM10000 and PH2. The originality of this survey proved that DEANN was better than the other conventional classifiers regarding recognition of correctness. The outcomes proved that this suggested method recognized SC efficiently and gave 97 percent correctness.**Hasan, S. N., et al., (2019) [14]** proposed the automated segmentation of dermoscopic pictures of the skin lesion field. The main goal was to develop a system for segmenting the lesion that was accurate, efficient, strong, and automatic, allowing for a more precise categorization of the lesion at the time of early SC detection. Pre-processing and picture segmentation were the two parts of this approach. Image processing techniques such as Enhancement, Filtering, and Restoration were used to provide images free of artifacts such as hair and ruler marks for the first phase. This model's next phase is critical since it changed a U-Net framework and introduced a 46-Layered U-Net framework for obtaining an efficient lesion segmentation rate. The tests were carried out on two different U-Net frameworks (U-Net 32 and 46). Using the ISIC2018 database of 1815 photos and estimating it on 779 validation databases, the framework U-Net 46 attained 93 percentAcc, 91 percentSN, and 97 percentSP.**Albahar, M. A., et al., (2019) [15]** presented a novel estimation design that categorized SC's as mild or harmful based on a new regularized method. Therefore, it is a binary categorizer differentiatingbetween mild or harmful lesions. The suggested method attained 97.5 percent accuracy, proving its supremacy compared to other techniques. The accomplishment of CNN (convolutional neural network) regarding AUC-ROC with a new regularize was examined under different application circumstances. The AUC parameter attained for nevus beside melanoma lesion, seborrheic keratosis versus basal cell carcinoma lesion, seborrheic keratosis versus melanoma lesion, solar lentigo versus melanoma lesion was 0.7, 0.9, 0.8, and 0.8, correspondingly. Table 1 discusses various ML and DL methods used to detect skin cancer disease. Analysis of several methods to extract the reliable features and improve the performance metrics like accuracy, precision, etc.

Table 1: Analysis with various Detection Methods

Author Name	Techniques	Dataset	Parameter	Future Scope
Nawar A et al., 2021 [9]	Color Segmentation GLCM features Statistical features K-means Clustering SVM Classifier	500*400 image pixels	Accuracy (Acc) SN SP Precision FPR FNR	It will propose work with more skin Maladies and the acc rate.

Ahmed et al., 2021 [10]	CNN-SVM-MAA	3000 images Beni-Suef University Hospital	Detection and recognization rate	It will help detect SD in rural parts of India, where there is a significant lack of standard medical facilities.
Srinivasu, P. N et al., 2021 [11]	MobileNet is a CNN-based method MobileNet V2 and MobileNet V2 with LSTM	HAM10000	SN SP Accuracy JSI and MCC	It will perform to study the FE (feature extraction) actionsbased on biomarkers.
Naeem, A et al., 2020 [12]	CNN	Non-published dataset	SN SP Pre Accuracy AUC	-
Kumar, M et al., 2020 [13]	LBP DE-ANN classifier	HAM10000 and PH2	Acc SP SN	It will extend toanother category of SC diseases using DL-based methods.
Hasan, S. N et al., 2019 [14]	U-net 32 and 46 architectures	448*488 pixels (ISIC 2018)	JSI and SP	-
Albahar, M. A. 2019 [15]	Deep-CNN	Benchmark dataset 600*600 pixels	Acc AUC SN SP	It will produce better results than the existing survey.

Abbrevations: GLCM (Gray Level Co-occurance Matrix); SVM Classifier (Support Vector Machine); SN (Sensitivity); SP(Specifity); Pre (precision); FPR(False positive rate); FNR (false negative rate);CNN-SVM-MAA (Convolution Neural Network-Support Vector Machine-Mobile Android Application); LSTM (long short term memory); JSI (Jaccard Similarity Index), and MCC (Mathew Coefficient Correlation); LBP(linear binary pattern);DE-ANN classifier (Differential evolution-Artificial Neural Network); Deep-CNN (Deep-convolution-neural-network); DI (dermoscope image);FTNN (Fine-tuned neural networks);LSIR(Long-scale image recognition);VGG (Visual Geometry Group); DCNN (Deep-CNN);DL (Deep learning).

3. Proposed Methodology

The research method is introduced in several stages to attain reliable outcomes of SC detection. The limitations of SL detection with DI. It is not easy for skin experts to verify whether the specific skin is cancer /non-cancer. To verify outcomes for SC, they are required to carry out particular clinical tests that become costly and time-consuming with minimum precision for the patients. As an outcome, an exact automatic segmentation is needed, which is the centralized and minimum cost. Hence, a new method is defined with a maximum accuracy rate. The KPCA is a feature extraction approach and occasionally leads to imprecise outcomes. So, in the research method, feature extraction, and hybrid featured-optimized MSVM classification. It enhances the Acc. rate of the method if the feature sets are precisely assessed, then the possibility of getting an exact outcome grows and optimizes the computation cost.

In the initial phase of the research model, preprocessing was applied to create the input DI smooth, filtered, and noise-free. So, a median filter (MF) is used to alter the unwanted noise. The research flowchart is defined in figure 1. In the next phase, IS (image segmentation) divides a DI into disjoint fields depending on various metrics like; color, surface, etc. To segment homogenous clusters, the FCM method is applied. FCM method used by providing membership to the data-point (DP) depends on the distance between the cluster-center“c” and DP. If more information is near to “c” then more is the DP membership for that cluster. After that segmentation, introduce and assess the feature sets of image text in reply to non-deterministic feature sets, the relationship between the GLs (graylevels) of the cancer image is needed. The properties of a DI are calculated using co-occurrence matrics based on distance and angular relationsamong image pixels. A matrix that defines row and column signifies the gray-levels “g” in the image m*n nearest pixes and intensity I(m,n). After the fetch the properties using the KPCA method. This method has extricated the reliable feature set. After the feature extraction process has been implemented,the GWO optimization method. This optimization process has selected reliable feature sets in the form of matrices. To assess more reliable feature sets in the research method, the GWO is developed as a classification method. The research method using Hybrid Featured- Optimized MSVM classifier is used to classify or detect the images of cancer and non-cancer. The scheming procedure of MSVM is challenging and has several problems. These problems might be regarded asselecting an influentialnetwork, LR (learning rate), class division, train, test,etc. Thus, Hybrid Featured-Optimized MSVM is trained to utilize grey wolf optimization to improve accuracy and identify such problems in the research work.

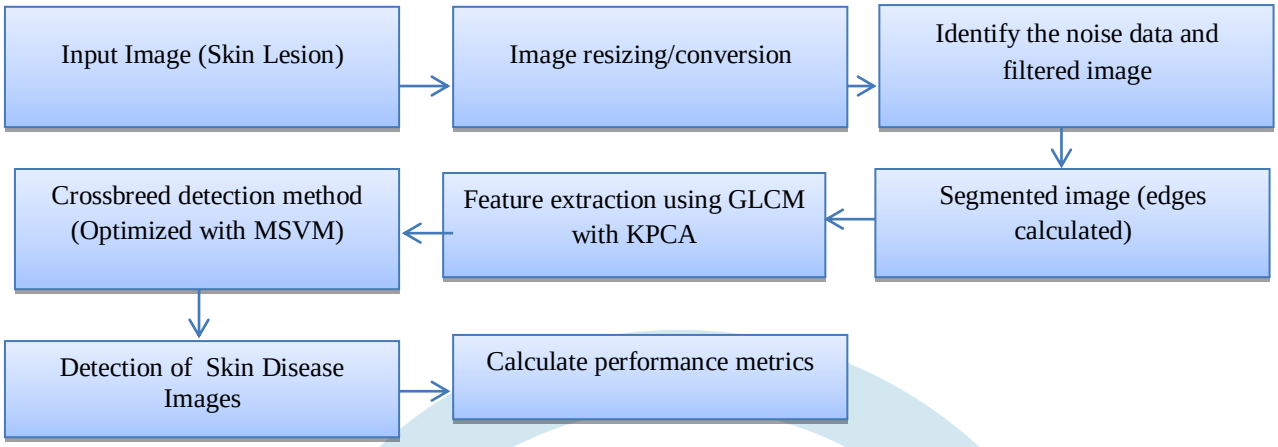


Fig 1. Proposed Flow Chart

4. Experiment Results and Discussions

The proposed database used in this work is HAM10000 [16] [17]. For the proposed motive, these databases are freely available and have a repository of 10,015 and 200 DIs, respectively. The calculation of the research methods is completed using HAM10000 databases shown in table 2. Using these databases, 120 cancer and non-cancer images are collected to evaluate this simulation.

Table 2. HAM10000 Data with Different pre-processing Steps

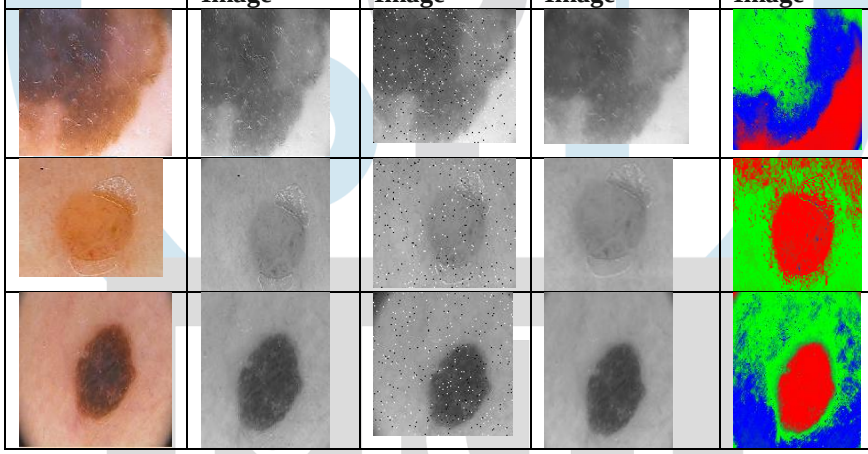
Input Image	**Conversion Image**	**Attacked Image**	**Smooth Image**	**Segmented Image**
				

Table 2 shows that the input image is a color image; it converts it to a grayscale image. It finds the attack skin cancer image and filters the uploaded image. After that, it finds the segmented image based on colored features.

These DIs the research Hybrid-feature-optimized MSVM detected cancer and non-cancer. The simulation setup has done using by MATLAB tool. Several metrics are utilized to set up the simulation with hybrid-featured-optimize MSVM model. Using the research simulation setup, the research approach can get an accuracy of 98.0 percent. The proposed results verify that the novel introduced approach is efficient for three categories of SC detection with a maximum accuracy rate. Table 3 defines the extracted feature set values and classifier results. An image sample outcome is defined in Table 3 for randomly selected three images (1-1 image from three categories of DI). It is considered that the defined technique precisely extracted features of the skin cancer image and assessed feature sets are further utilized for training the Hybrid-feature-optimize MSVM classifier.

Table 3. Research method outcomes for three different categories of cancer and non-cancer DI images.

Category of DI image	Contrast	Energy	Correlation	Homogeneity	Class
Common Nervous	0.12176	0.25372	0.95545	0.93931	0
Atypical Nevus	0.095426	0.21431	0.98554	0.95352	1
Melanoma	0.11463	0.15011	0.98014	0.95077	2

To study the FE (feature extraction) and classification outcomes, different performance metrics such as; accuracy, SP, SN, etc. These performance metrics are defined as below:

$$acc = \frac{(t_p+t_n)}{(t_p+t_n+f_p+f_n)} \quad \text{…….. (i)}$$

$$SN = \frac{(t_p)}{(t_p+f_n)} \quad \text{…….. (ii)}$$

$$SP = \frac{(t_n)}{(t_n+f_p)} \quad \text{…….....(iii)}$$

Here eq (i),(ii), and (iii) shows the Tp = true_positive; Tn = true_negative; Fp = false_positive; Fn = false_negative. The proposed method performance is calculated using Tp and Fps.

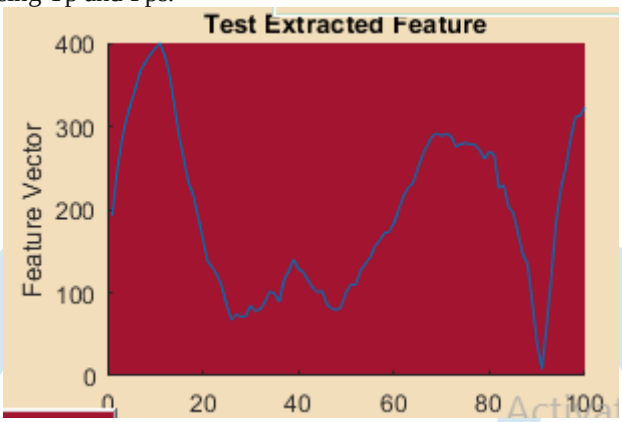


Fig 2. Test extracted feature sets

Fig 2 defines the test feature sets with GLCM and KPCA feature extraction methods. KPCA algorithm has been used to reduce the dimensionality of uploaded skin cancer image features. The proposed method performance metrics for the hybrid-featured-optimize MSVM classifier method in terms of acc, SP, SN [18], are defined in fig 3 and table 4.

Table 4 Proposed Parameters

Parameters	**Acc (%)**	**SP (%)**	**SN (%)**
Values	98.0	0.972 ~ 97.2	0.962

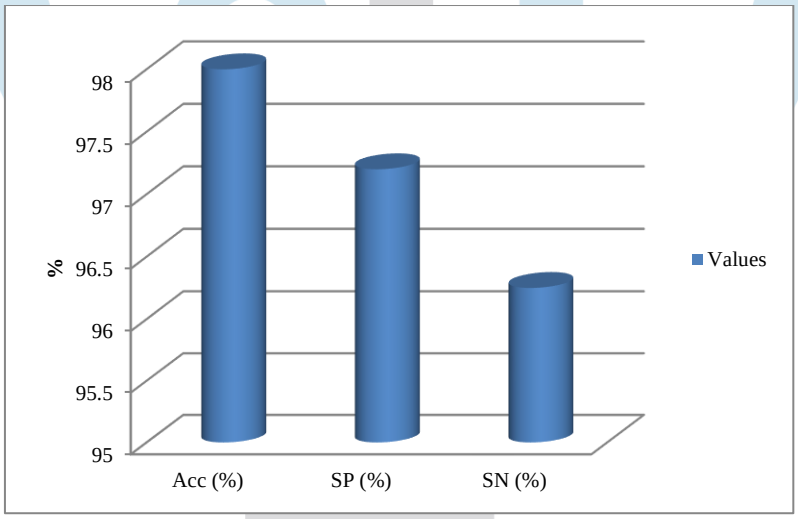


Fig 3. Proposed Parameters with Hybrid-feature-optimize MSVM classifier

Table 5. Comparative Analysis with proposed and existing models

Methods/Parameters	**Accuracy (%)**
SVM	86
ANN	88
GA-ANN	94.6
DE-ANN	97.7
Hybrid-featured-Optimize-MSVM	98.0

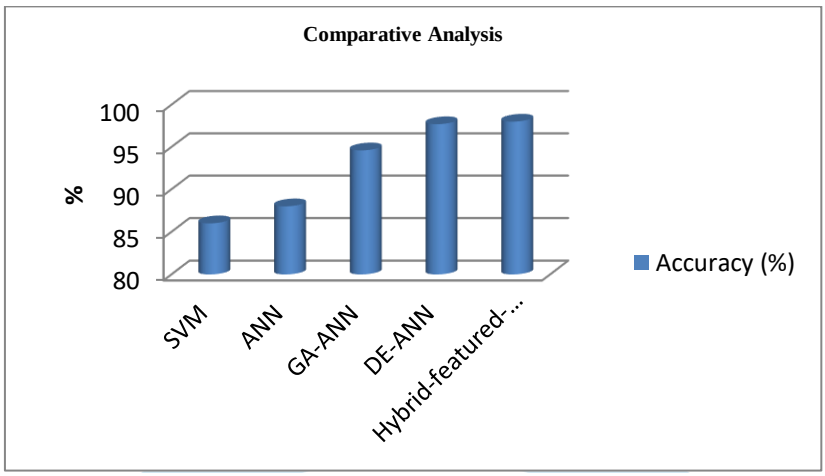


Fig 4. Comparative analysis with proposed and existing Models

Fig 4 and Table 5 show the accuracy classification is about 98.0 percent for the research technique compared with the SVM value of 86 percent. The ANN value is 88 percent, the GA-ANN method is 94.6 percent, the DE-ANN method is 97.7 percent, and the proposed method value is 98.0 percent. Also, in the research, hybrid-featured-optimized MSVM is implemented as a prediction to classify the DIs into cancer and non-cancer forms of outcomes. The proposed model is that this is considered one of the most reliable and efficient ML-based methods for a high accuracy rate. Fig 4 defines a comparison analysis between proposed and existing models such as GA-ANN [20], ANN [19], SVM [18], and DE-ANN [13] model.

5. Conclusion and Future Improvements

It concluded that the improved SD is classified as an earlystage. The presented work is implemented to classify or detect the early signs and symptoms of three different categories of SC precisely using DIP (digital image processing) and computer-based methods. This proposed method has implemented the classification of SC using hybrid-featured-optimize MSVM model. All three categories of SC dermoscopic image features are extracted using KPCA, GLCM, and RGB color-space. The feature Extraction method has extracted the features which have a maximum contribution to the skin disease classification model. The outcome defines that the research method is broad-minded towards image segmentation using the clustering method or color-space model. Utilizing the HAM10000 DI database, the accuracy rate of the method is about 98.0 percent. Generally, the research architecture is compared with other ML-based methods and accomplishes what the researched one is well. The proposed consequences are also improved as linked with other existing techniques. In the upcoming work,the research may be improved to assess the correlation betweenskin itching due to outside metrics like; sunburn. This proposed work may be improved or prolonged for other categories of SC diseases using DL methods such as LSTM, RNN, GNN, etc.

References

1. Das, K., Cockerell, C. J., Patil, A., Pietkiewicz, P., Giulini, M., Grabbe, S., & Goldust, M. (2021). Machine Learning and Its Application in Skin Cancer. *International journal of environmental research and public health*, *18*(24), 13409.
2. Ferlay, J., Colombet, M., Soerjomataram, I., Parkin, D. M., Piñeros, M., Znaor, A., & Bray, F. (2021). Cancer statistics for the year 2020: An overview. *International Journal of Cancer*, *149*(4), 778-789.
3. Apalla, Z., Nashan, D., Weller, R. B., & Castellsagué, X. (2017). Skin cancer: epidemiology, disease burden, pathophysiology, diagnosis, and therapeutic approaches. *Dermatology and therapy*, *7*(1), 5-19.
4. Davis, L. E., Shalin, S. C., & Tackett, A. J. (2019). Current state of melanoma diagnosis and treatment. *Cancer biology & therapy*, *20*(11), 1366-1379.
5. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *nature*, *542*(7639), 115-118.
6. Bohr, A., & Memarzadeh, K. (2020). The rise of artificial intelligence in healthcare applications. In *Artificial Intelligence in healthcare* (pp. 25-60). Academic Press.
7. Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific data*, *5*(1), 1-9.
8. Mendonça, T., Ferreira, P. M., Marques, J. S., Marcal, A. R., & Rozeira, J. (2013, July). PH 2-A dermoscopic image database for research and benchmarking. In *2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC)* (pp. 5437-5440). IEEE.
9. Nawar, A., Sabuz, N. K., Siddiquee, S. M. T., Rabbani, M., Biswas, A. A., & Majumder, A. (2021, March). Skin disease recognition: A machine vision based approach. In *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)* (Vol. 1, pp. 1029-1034). IEEE.
10. Elngar, A. A., Kumar, R., Hayat, A., & Churi, P. (2021, August). Intelligent System for Skin Disease Prediction using Machine Learning. In *Journal of Physics: Conference Series* (Vol. 1998, No. 1, p. 012037). IOP Publishing.
11. Srinivasu, P. N., SivaSai, J. G., Ijaz, M. F., Bhoi, A. K., Kim, W., & Kang, J. J. (2021). Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM. *Sensors*, *21*(8), 2852.

12. Naeem, A., Farooq, M. S., Khelifi, A., & Abid, A. (2020). Malignant melanoma classification using deep learning: datasets, performance measurements, challenges and opportunities. *IEEE Access*, *8*, 110575-110597.
13. Kumar, M., Alshehri, M., AlGhamdi, R., Sharma, P., & Deep, V. (2020). A de-ann inspired skin cancer detection approach using fuzzy c-means clustering. *Mobile Networks and Applications*, *25*, 1319-1329.
14. Hasan, S. N., Gezer, M., Azeez, R. A., & Gülseçen, S. (2019, October). Skin lesion segmentation by using deep learning techniques. In *2019 Medical Technologies Congress (TIPTEKNO)* (pp. 1-4). IEEE.
15. Albahar, M. A. (2019). Skin lesion classification using convolutional neural network with novel regularizer. *IEEE Access*, *7*, 38306-38313.
16. K. Inc. (2019) Skin Cancer MNIST: HAM10000. Available:https://www.kaggle.com/kmader/skin-cancer-mnist-ham10000/version/2. Accessed 2022.
17. ADDI Project PH2 database, available: https://www.fc.up.pt/addi/ph2%20database.html. Accessed May 2022.
18. Chang, C. C., & Lin, C. J. (2011). LIBSVM: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, *2*(3), 1-27.
19. Amato, F., López, A., Peña-Méndez, E. M., Vaňhara, P., Hampl, A., & Havel, J. (2013). Artificial neural networks in medical diagnosis. *Journal of applied biomedicine*, *11*(2), 47-58.
20. Ahmad, F., Mat-Isa, N. A., Hussain, Z., Boudville, R., & Osman, M. K. (2010, July). Genetic algorithm-artificial neural network (GA-ANN) hybrid intelligence for cancer diagnosis. In *2010 2nd international conference on computational intelligence, communication systems and networks* (pp. 78-83). IEEE.