NeuroDerm: Intelligent Skin Disorder Diagnosis Using ML

[1] Ms.YASHODHA P.G [2] RAKESH N R
[1] Assiatant professor, Department of MCA,BIET,Davanagere
[2] Student,Department of Master of Computer Applications, BIET,Davanagere

ABSTRACT

Skin conditions are the most common in health problems throughout the world. These illnesses have hidden risks that can be lead to physical discomfort and occasionally even depression. Skin cancer may even result from severe cases. One of this most of the challenging tasks in medical image processing is the diagnosis of skin disorders from clinical photographs. When completed by medical professionals by hand, it takes a lot of this time and is subjective. Thus, in way to expedite the process in developing the treatment plans for the patients and the dermatologists, an automatic method for predicting skin diseases is must required. In this our work, we suggest a method for digital hair removal that combines an inpainting algorithm with morphological filtering, such as Black-Hat transformation, and then uses Gaussian filtering to de-noise or de-blur the images. To separate the impacted lesions, we additionally employ the automated Grabcut segmentation method. In the way to identify patterns in the skin photos,

1. INTRODUCTION

The Human body comprises various organs, and the skin is the largest, covering the entire body. Any disorder affecting the skin is termed a skin disease. Skin diseases are among these most contagious worldwide. According to WHO, 17857 people in India died from the skin cancer in 2020. Globally, over 14 million cases were diagnosed, and 9.6 million deaths occurred in 2018. Skin disease often involves a change in color or texture of the skin. Causes include viruses, bacteria, allergies, or fungal infections. Genetic factors were can also be lead to skin disorders. Skin diseases typically affect the thin outer layer, the epidermis, and to can be seen by the human eye, causing psychological depression and physical injuries.

There are various types of the skin lesions, including The Actinic keratosis (AK), The Basal cell carcinoma (BCC), The Benign keratosis (BKL), Dermatofibroma (DF), The Melanoma (MEL), The Melanocytic nevus (NV),The Squamous cell carcinoma (SCC), and The Vascular lesion (VASC). These lesions differ in symptoms and severity. Some are permanent, some are temporary, and they can be either painless or painful. Melanoma in which this the most deadly and dangerous among these skin diseases. However, about 95% of the skin disease patients can recover if the condition has been identified early. An automatic computer-aided system can help accurately classify skin diseases. There is a significant gap between dermatologists and patients with skin diseases, as many people are unaware of the types, symptoms, and stages of these conditions. Sometimes, signs take a long time to appear, necessitating early and quick detection. However, diagnosing the skin diseases correctly to identify particular type and stage can be difficult and expensive. Machine learning-based automatic computer-aided systems have made it possible to detect the skin disease types more accurately and quickly.

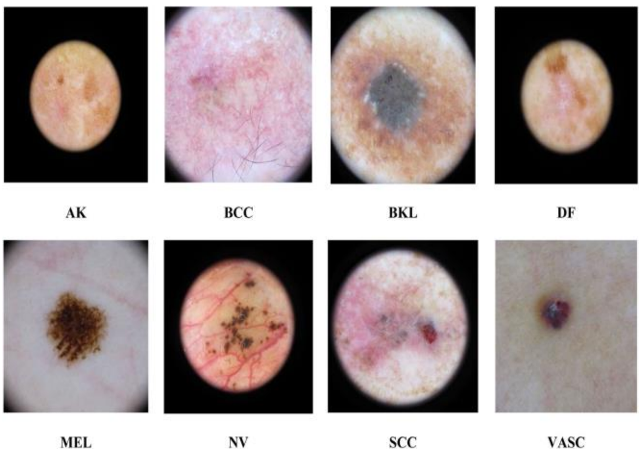


fig. 1. Sample of the Skin Disease Images which collected from in ISIC 2019 Challenge Dataset.

2. LITERATURE SURVEY

Numerous scholars have to put forth the different methods for categorizing skin conditions. Based on datasets, feature on extraction methods, feature on selection methods, and classification models, related works are grouped. In way of examine the methods and resources employed in earlier studies and to pinpoint research needs, this section examines a total number of pertinent research publications.

Jagdish et al. [9] who introduced a new skin disease of detection model using the image processing methods. They applied fuzzy clustering on 50 sample images with KNN and SVM classification algorithms alongside wavelet analysis. Their results demonstrated that the K-Nearest Neighbor algorithm outperformed in the Support-Vector-Machine (SVM) technique with an accuracy of 91.2%, successfully identifying type of the skin disease, albeit with a limited dataset which containing only the two classes (basal and squamous disease).

Naeem et al. [10] proposed on a skin cancer of prediction model utilizing image processing strategies and support-vector-machines (SVMs). They employed various preprocessing of techniques for the noise removal and the image enhancement, along with the GLCM method for extracting several features from which the image. The classifier then categorized the images as harmful or harmless.

Bandyopadhyay et al. [11] introduced a model that is to combines the deep learning (DL) and the machine learning (ML) techniques. They utilized deep neural networks like the Alexnet, Googlenet, Resnet50, and VGG16 for feature selection and Support-Vector-Machine, Decision tree and Ensemble boosting Adaboost classifiers for classification. A comparative the study was conducted to determine which the best prediction model.

Kalaivani et al. [12] proposed the novel method that integrates two separate data mining approaches into a single unit, along with an ensemble approach that merges both data mining techniques into a single group. They applied the ensemble deep learning in technique to the ISIC2019 dataset, categorizing skin disorders into seven categories. The ensemble technique demonstrated more accurate and effective predictions of the skin diseases.

Aldera etal. [13] who presented the skin disease diagnosis model which capable of diagnosing acne, cherry angioma, melanoma, and psoriasis using affected skin images. They applied Otsu's method for image segmentation and Gabor, Entropy, and Sobel techniques for feature extraction. Support-Vector-Machine (SVM), Random-Forest (RF), and K-Nearest Neighbor (K-NN) classifiers were employed for classification, achieving accuracies of 90.7%, 84.2%, and 67.1%, respectively.

Kshirsagar et al. [14] proposed a the skin disease classification of system the using that the where MobileNetV2 and the LSTM. Their primary focus which was on the accuracy in the skin disease forecasting, ensuring efficient storage of complete state of information for the precise forecasts. Comparisons with traditional models which such as CNN and FTNN showed in that the proposed which model which outperformed.

Hatem [15] proposed the system which to identify the skin lesions and to classify them as the normal or the benign. They utilized a k-nearest-neighbor (KNN) classifier achieving an accuracy of 98% but only for the two classes.

Kethana et al. [16] introduced a model employing Convolutional-Neural-Network (CNN) for the skin disease classification. They utilized the dataset of the 10,015 images from which the ISIC 2019 dataset and achieved a classification accuracy of 92% for Melanoma, Nevus, and Seborrheic Keratosis.

Yao et al. [17] proposed a novel single-model-based approach for classifying skin lesions on small and imbalanced datasets. They trained various Deep Convolutional Neural Networks (DCNNs) on small and imbalanced datasets and implemented regularization techniques such as DropOut and DropBlock to reduce overfitting. They achieved high classification performance using a Modified RandAugment augmentation strategy, a Multi-Weighted New Loss (MWNL) function, and an end-to-end cumulative learning strategy (CLS).

Padmavathi et al. [18] presented an automated in which the skin lesion of classification approach using a pre-trained and fine-tuned deep learning network. They which evaluated the performance using various quantitative criteria and compared different transfer learning methods.

Maduranga etal. [19] developed an artificial intelligence-based mobile application for the skin disease the detection. They employed Convolutional-Neural-Networks (CNNs) on this the HAM10000 in which dataset and built a mobile application using MobileNet with transfer learning, achieving an this accuracy where of 85%.

Jain etal. [20] proposed an optimal probability-based deep this neural network (OP-DNN) for the skin disease diagnosis. They which extracted features from images and trained the OP-DNN.

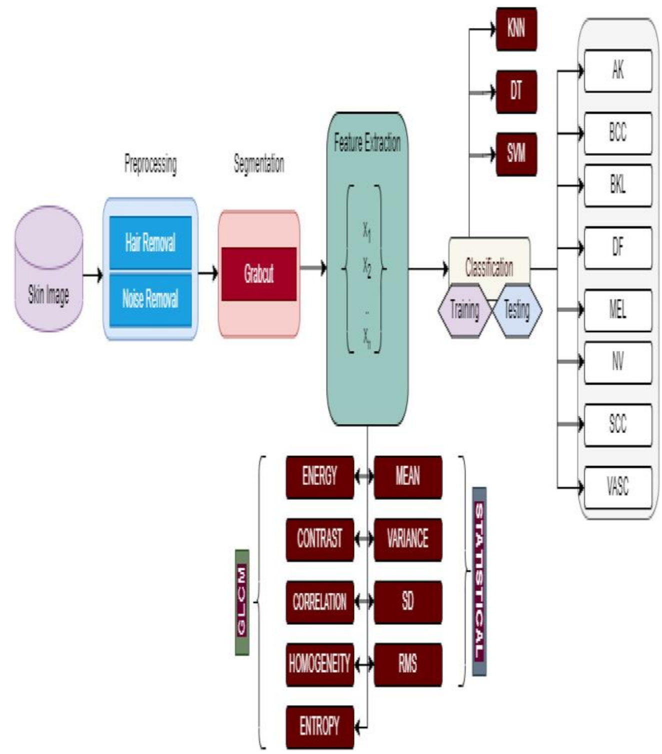


Fig 2. A Generalized Block Diagram of the Skin Diseases Classification system

3. Methodology

This described and discussed the proposed approaches for classifying the skin diseases. which whole process comprises the following parts: The first step is image preprocessing, the second step is image segmentation, the third step is feature-extraction, and the final step is classification. Fig. 2 depicts a high-level overview of where proposed strategy.

3.1 Image Preprocessing.

Image preprocessing that involves converting images into a more suitable and usable form. Skin images often contain unwanted elements like hair, noise, or distortion, which can affect system performance. Preprocessing aims to enhance image quality, reduce complexity, and improve accuracy. The preprocessing steps include:

3.2 Image Resizing

Images of varying sizes may result in differences in feature extraction. Resizing standardizes image dimensions, ensuring consistency and facilitating processing. In our work, all the input images were resized where to 512 × 512 pixels for improved system performance.

[Fig. 3: Original image (a) and scaled image (b)]

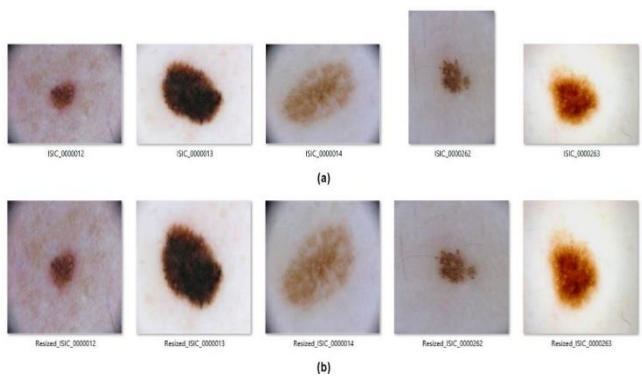


Fig. 3. Figure which shows image preprocessing (a) Before Resizing (b) After resizing

4. Feature Extraction

Feature-extractions is the essential for studying and discovering the underlying of relationships between the various objects. The image categorization, prediction, and recommendation algorithms cannot comprehend of the images directly. As the result, the feature which extraction is much necessary to convert them into usable forms. The dermoscopic image has the various of the characteristics that which are to be utilized to describe the image. However, not which all characteristics are which will apply the categorization of skin disease. As the result, the classifier becomes complex and takes more to the computational effort in several irrelevant features, reducing on the classification accuracy. In those skin cancer pictures, the best features that must reflect the characteristics of the areas. As a result, a sufficient number of features distinguish the images which as accurately is feasible. The best method to handle the region in isolation is to use the segmented lesion images to extract many features for this task. In these work, we need used the GLCM features as texture features and several statistical features as color features to the determine the skin disease type.

4.1 GLCM Features

First, the Gray Level Co-occurrence Matrix (GLCM) computes each image. The contrast, energy, entropy, correlation, and homogeneity features were which then are calculated using those matrix. Table 1 shows the extracted GLCM features, along with their descriptions and formulas.

4.2 Statistical Features

The RGB (red, green, and blue) color spaces are explored for each of image, and various statistical-features are extracted. The statistical features extracted in study included the mean, variance, standard deviation, and root mean square. Table 2 which shows extracted-features, along its their descriptions and formulas. The extracted feature values for some sample images are shown in Table 3.

5. Classification

The final task in the work is classification. It is the process of dividing a set of data into different categories. This study predicted the type of skin disease using features extracted from images. Classification methods vary by application and dataset type. In this study, we used three different classifiers to categorize skin diseases: Support-Vector-Machine (SVM), K-Nearest-Neighbor (KNN), and Decision Tree (DT). A Support-Vector-Machine (SVM) is a supervised classification technique for solving the classification and regression problems. It produces the most accurate results to than most other algorithms.

To classify eight skin diseases, we used multiclass SVM with two approaches: one-to-one and one-to-rest [44]. It uses the kernel technique to transform data, resulting in an optimal decision boundary

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 07 | July - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

between the possible outputs [45]. K-Nearest Neighbor is a nonparametric method that can need to applied both classification and regression problems. It predicts the range of unknown data points by using 'feature similarity'. In our study, the similarities between the new and existing cases are determined by calculating the maximum distance between to the two data points using the Euclidean distance [46]. The decision tree represents the knowledge in a tree structure to make it more understandable.

6. Dataset

6.1 Dataset Description

We used two well-known datasets: the "International Skin in Imaging Collaboration (ISIC) 2019" challenge dataset and "HAM1000

ISIC 2019 Challenge Dataset

An international database of dermatoscopic images is called the ISIC [47]. It has evolved to support both technical difficulties and clinical practice. We have exclusively utilized the ISIC 2019 challenge training dataset, which comprises 25331 dermoscopic images across across eight distinct classes: "Actinic keratosis (AK), the basal cell carcinoma (BCC), the benign keratosis (BKL), dermatofibroma (DF), the melanoma (MEL), the melanocytic nevus (NV), the squamous cell carcinoma (SCC), and vascu- lar lesion." The dataset includes the HAM10000 (Human-Against-Machine with 10000 training images).

6.2 Dataset Preparation

The ISIC 2019 challenge training dataset includes 25331 images divided into eight categories: "the basal cell carcinoma (BCC), the actinic keratosis (AK), the melanoma (MEL),the melanocytic nevus (NV),the benign keratosis (BKL), the dermatofibroma (DF), vascular lesion (VASC), and squamous cell carcinoma (SCC)." The dataset is imbalanced, with 867 images of class AK, 3323 images of class BCC, 2624 images of class BKL, 239 images of class DF, 4522 images of class MEL, 12875 images of class NV, 628 images of class SCC, and 253 images of class VASC. The HAM10000 dataset is also an imbalanced dataset. The AK class has 327 images, the BCC class has 514 images, the BKL class has 1099 images, the DF class has 115 images, the MEL class has 1113 images, and the NV class. The data distribution of each class for ISIC 2019 dataset is in Figure. 4. While training to the model with the imbalanced dataset, there is a bias that predicts the majority classes while ignoring the minority classes. As a consequence, the error minority classes benefit from increases while majority classes suffer from decreases. We used the data balancing Random Oversampling.method to resolve the imbalanced dataset in both datasets.

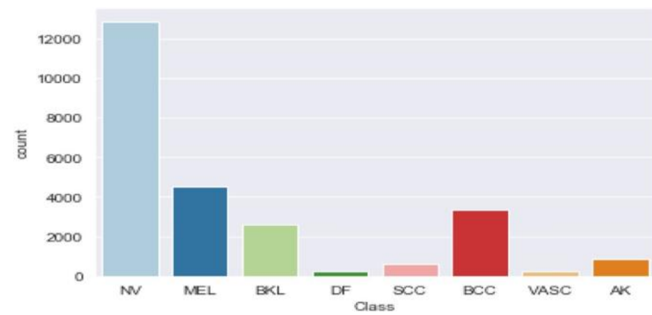


Fig 4.The distribution of the maximum number of the Images into eight different classes

6.2.1 Random Oversampling

Oversampling is a technique whch for randomly duplicating samples from minority classes. The smallest class is considered a minority class. The process is repeated until no class has a sample size less than the largest. Finally, sampling is completed once the balance has been achieved. We used the oversampling method for both the ISIC 2019 and HAM10000 datasets. Here, we only showed the distribution of the balanced ISIC 2019 dataset using the oversampling method shown in the Fig. 5

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 07 | July - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

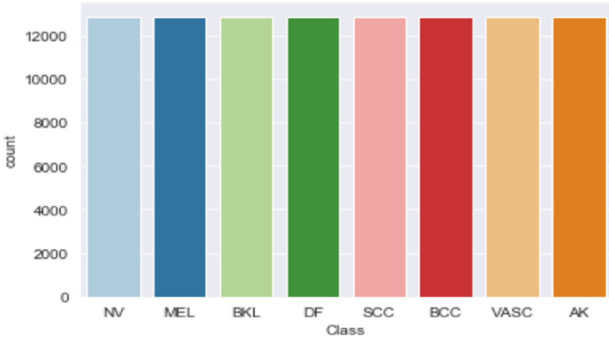


fig. 5. Distribution of data after random Oversampling

Performance metrics of this proposed models that's for balanced dataset.

7. Result

Dataset Model		Accuracy	Precision	Recall	F1-score	Log loss (%)
ISIC 2019	SVM	95.00	95.13	95.00	94.88	18.09
	KNN	94.00	93.88	93.88	93.38	25.49
	DT	93.00	93.13	93.00	92.50	24.49
HAM10000	SVM	97.00	97.71	97.57	97.43	11.37
	KNN	95.00	95.71	95.57	95.14	15.59
	DT	95.00	95.14	95.14	94.71	17.37

This section which presents our assessment of the suggested classification model's performance along with a comparison to the particular performance of other methods currently in use. We have tested with our model both which with the unbalanced datasets and after balancing the datasets because our datasets are unbalanced. improvement in sensitivity results using the SVM algorithm (95%) in comparison to the decision tree (93%) and even the KNN (94%) algorithm, indicating that its the model outperforms on the large dataset using the SVM algorithm, as shown in Fig. 11 and Table 7. Nonetheless, the average precision and average F1 score values obtained using the SVM algorithm (95.13% and 94.88% respectively) are also greater.. Table 7 also shows that the SVM classifier outperforms the three classifiers for both datasets, with an average accuracy of 95% and 97%, respectively. We also made tested our particular proposed model that the HAM10000 dataset and discovered that it performs well, has been shown in

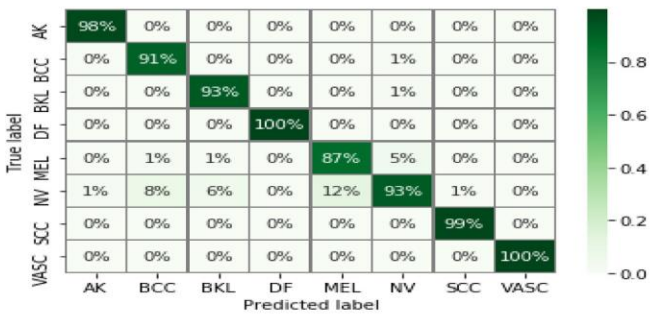
Furthermore, the confusion matrices presented in the Fig 6 and 7 demonstrate the major classification performance which are the investigated algorithms. In which the confusion matrix, that the value in each row represents the actual labels, while the value in each row represents the predicted labels. Thus, the cell value represents the percentage of prediction. At the given same time, it is the diagonal cell displays the highest level.

The confusion matrix predicts a minimum given error rate to for each class of skin diseases. We must used the confusion matrix to calculate the predicted accuracy, given precision, to recall, and f1-score to given evaluate the proposed model's performance.

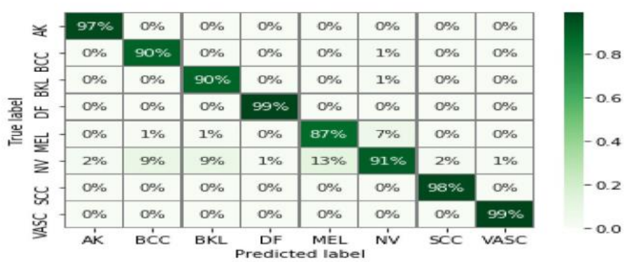
To solve our research problems, we also evaluated the classification The performance which of our given proposed models that for each given type of the skin disease is illustrated by the results of the ROC analysis shown in Figs. 9 (a, b, c) and 10 (a, b, c). Again, each classifier demonstrated similar behavior. We represented the type that disease in the ISIC 2019 dataset as

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 07 | July - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

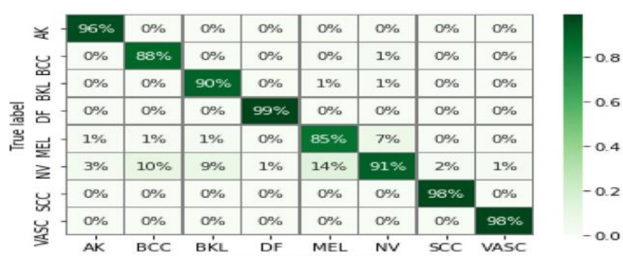
follows: Class 0: Actinic Keratosis (AK), class 1: the Basal Cell Carcinoma (BCC), class 2: Benign Keratosis (BKL), class 3: Dermatofibroma (DF), class 4: Melanoma (MEL), class 5: Melanocytic Nevus (NV), class 6:The Squamous Cell Carcinoma (SCC), class 7: Vascular. achieved the minimum AUC. The overall AUC of this system was 99% for the SVM classifier. The AUC shows that SVM outperforms DT and KNN in terms which detect.



(a)



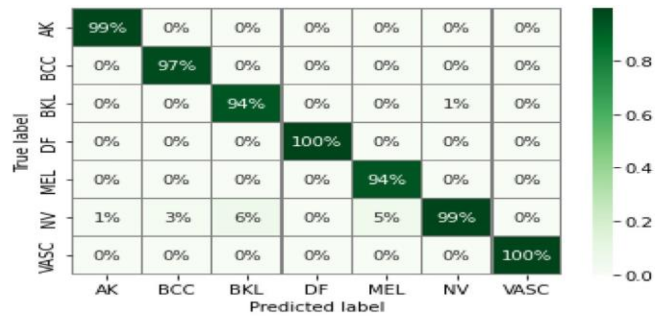
(b)



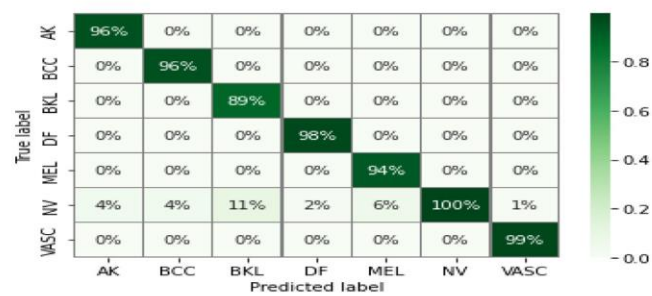
(c)

fig. 7. The given confusion matrix to the ISIC 2019 dataset of that the proposed system using the (a) SVM classifier, (b) KNN classifier, and (c) Decision Tree classifier, which represent the predicted values of the eight classes in percentage.

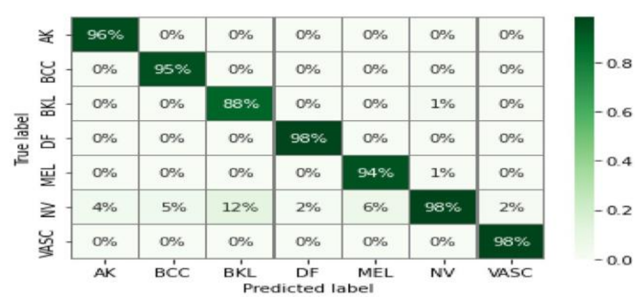
fig. 8. The given confusion matrix to the HAM10000 dataset of this proposed system using the (a) SVM classifier, (b) KNN classifier, and (c) Decision Tree classifier, which represent the predicted values of the eight classes in percentage



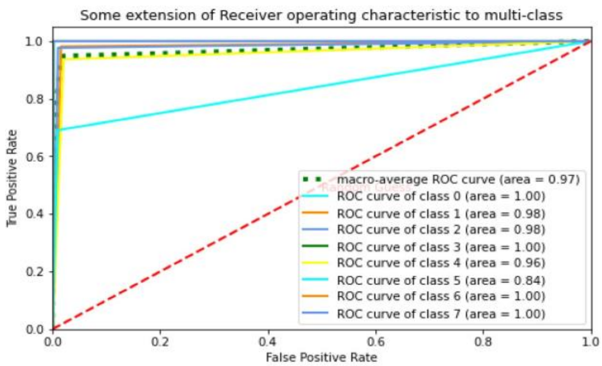
(a)



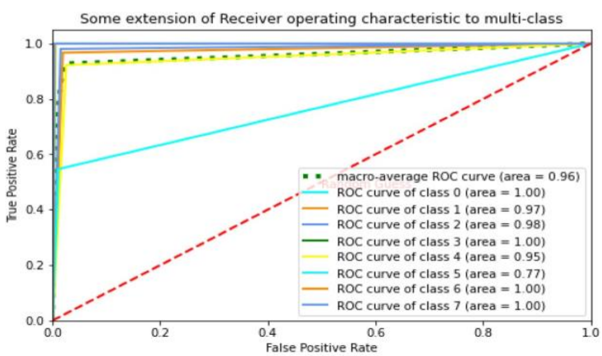
(b)



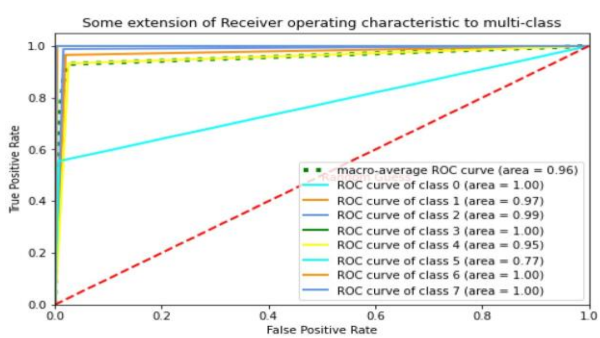
(c)



(a)

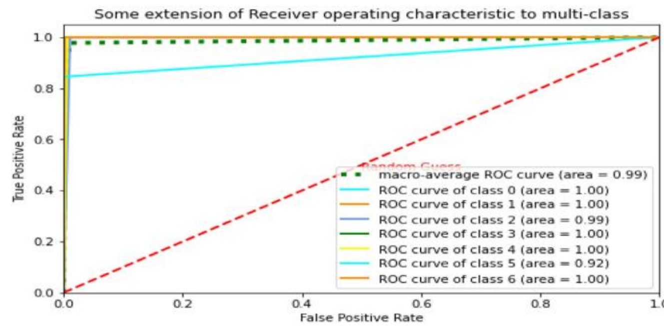


(b)

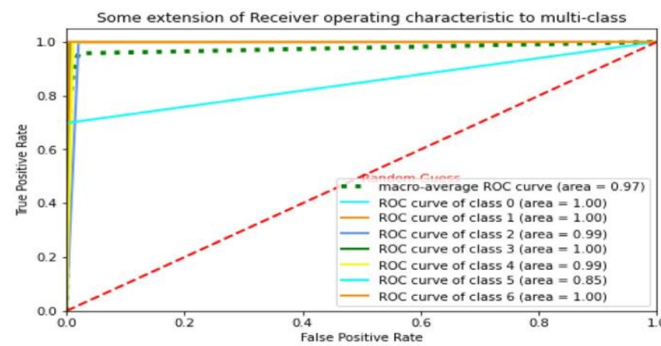


(c)

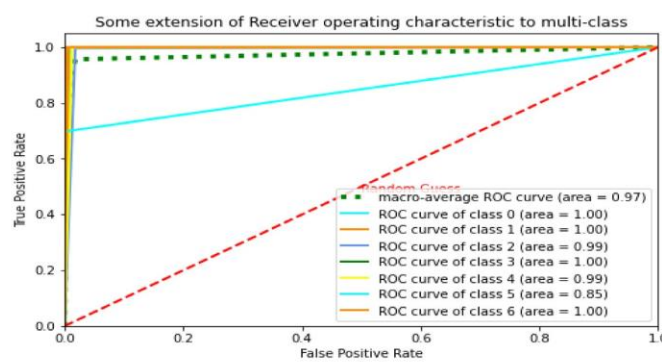
fig. 9. ROC performance curves for the ISIC 2019 dataset of this proposed classification model using (a) SVM classifier, (b) KNN classifier, and (c) Decision Tree Classifier. The curves close to the top left corner indicate the classification of accuracy of the each class in the SVM classifier. The area which Under the Curve (AUC) shows the model'smax ability to separate diseases.



(a)



(b)



(c)

fig. 16. ROC performance curves for the HAM10000 dataset of this proposed classification of model using ot the (a) SVM classifier, (b) KNN classifier, and (c) Decision Tree Classifier. The curves close to the top left corner indicate the most classification accuracy to each class in the SVM classifier. The area which Under the Curve (AUC) shows the given model's ability to separate diseases

8. Conclusion

Skin disease is now a global problem. People in many countries and regions who suffer from the various types of skin diseases. We can combat these diseases by developing different techniques and processes. We conducted this research in several phases. We can apply this model to other skin disease classification tasks. However, there still room for improvement in classification performance.

We used an automatic segmentation of algorithm, which given does not always identify the skin lesion correctly. As a result, misclassification occurs, which is one of our study's weaknesses. Future research will focus on real-time skin disease detection, using more efficient segmentation & classification techniques like ensemble learning and deep learning. Furthermore, we believe it will improve the speed and the accuracy of the image classification and the object detection algorithms. We hope it will help patients detect diseases early and maintain good skin health.

A customized digital hair removal technique was used, removing hairs with Morphological Black-Hat Transformation and blurring the images which a Gaussian Filter. Following that, we used an automatic Grabcut segmentation to detect to the skin lesion, which accurately segmented and identified the disease region. Finally, we extracted the GLCM and some statistical features and fed them into the SVM, KNN, and DT classifiers to determine to the type of skin disease. We used that two publicly available on benchmark datasets: the ISIC 2019 challenge and HAM10000. Because these datasets are somewhat imbalanced, we performed data balancing using the random over sampling technique. We achieved in an average given accuracy of 95%, 94%, and 93% for the ISIC 2019 dataset using SVM, KNN, and DT classifiers, respectively. Similarly, we achieved 97%, 95%, and 95% accuracy on which the HAM10000 dataset with SVM, KNN, and DT classifiers, respectively. It shows that this our model which performs better with the HAM10000 dataset than with the ISIC2019 dataset. We also discovered that this our model which performs exceptionally well for balanced data. Our model outperforms some of which most advanced methods for skin disease classification.

9. References

[1] Anatomy of skin, Stanford children's health, 2021, [Online]. Available: https://www.stanfordchildrens.org/en/topic/default?id=anatomy-of-the-skin-85-P01336

[2] M.W. Greaves, Skin disease, Britannica, 29, 2020, [Online]. Available: https: //www.britannica.com/science/human-skin-disease.

[3] India: Skin disease, 2018, [Online]. Available https://www.worldlifeexpectancy.com/india-skin-disease.

[4] Cancer, world health organization(W.H.O), 21, 2021, [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cancer.

[5] Md. Al Mamun, Mohammad Shorif Uddin, A survey on the skin disease detection system, Int. J. Healthc. Inform. Syst. Inform. 16 (4) (2021) 1–17.

[6] C.N.Vasconcenlos, B.N. Vasconcenlos, Experiment using the deep learning for dermoscopy of image analysis, of Pattern Re cognit. Lett. 139 (2020) 95–103.

[7] U.-O. Dorj, K.K. Lee, J.Y. Choi and. M. Lee, The skin cancer classification using deep convolutional neural networks, Multimedia of Tools Appl. 77 (2018) 9909–9924.

[8] M. Taufiq, N. Hameed, A. Anjum, F. Hameed, m-Skin Doctor: That A Mobile Enabled System for the Early Melanoma of Skin Cancer Detection Using the Support

Vector Machine, in: eHealth 360°. International Summit on eHealth, 2017, pp. 468–475.

[9] Jagdis etal., J.A.D.L. Cruz-Vargas, M.E.R. Camacho, Advance study of skin diseases detection using image processing methods, Nat. Volatiles Essent. Oils J. 9 (1) (2022) 997–1007.

[10] Z. Naeem, G. Zia, Z. Bukhari, A healthcare model to predict skin cancer using deep extreme machine, J. NCBAE 1 (2) (2022) 23–30.

[11] S.K. Bandyopadhyay, P. Bose, A. Bhaumik, S. Poddar,ML and deep learning integration for skin diseases prediction, Int. J. Eng. Trends of Technol. 70 (2) (2022) 11–18.

[12] A. Kalaivani, S. Karpagavalli, Detection and the classification of the skin diseases with ensembles of the deep learning networks in medical imaging, Int. J. Health Sci. 6 (S1) (2022) 13624–13637.

[13] S.A. AlDera, M.T.B. Othman, A given model for the classification and diagnosis of skin disease using ML and image processing techniques, Int. J. Adv. Comput. Sci. Appl. 13 (5) (2022).

[14] P.R. Kshirsagar, H. Manoharan, S. Shitharth, A.M. Alshareef, N. Albishry, P.K. Balachandran, Deep learning which approaches for the prognosis of automated of skin disease, Life 2022 12 (426) (2022).

[15] M.Q. Hatem, Skin lesion of classification system which using a Knearest neighbor algorithm, in: Visual of Computing for the Industry, and Art. Vol. 5, (7) 2022.

[16] K.S. A, M.S. B, Melanoma disease of detection and the classification of using deep learning, Int. J. Res. Appl. Sci. Eng. Technol. 10 (7) (2022).

[17] P. Yao, Single model deep learning on imbalanced small datasets for the skin lesion of classification, IEEE Trans. Med. Imaging 41 (5) (2022) 1242–1254.

[18] K. Padmavathi, H. Neelam, M.P.K. Reddy, P. Yadlapalli, K.S. Veerella, K. Pam- pari, Melanoma detection of using the deep learning, in: 2022 International Conference on Computer Communication and Informatics, ICCCI, 2022.

[19] M. Maduranga, D. Nandasena, Mobile-based skin disease has diagnosis system of using the convolutional neural networks (CNN), I.J. Image Graphics Signal Process. 3 (2022) 47–57.

[20] A. Jain, A.C.S. Rao, P.K. Jain, A. Abraham, Multi-type skin diseases classification using OP-DNN based feature extraction approach, Multimedia Tools Appl. 81 (2022) 6451–6476.