Title: Skin Disease Classification Model.

I. Introduction

Globally, there is an estimate of 3 billion people lacking access to medical care for skin disease (Daneshjou et al., 2022). Skin diseases adversely impacts patient's health and psycho-social life if no early diagnosis and treatment of the disease was done. Skin cancer like melanoma cancer is a form of fast spreading malignant skin disease, if diagnosed in primary stage and received timely treatment the patient stands a higher chance of recovery, however, if detected in advance stage will result in significant high mortality (Salem Ghahfarrokhi et al., 2022).

To meet the demand for accessible and fast diagnosis of skin disease, there is growing development of computer-aided diagnosis (CAD) systems that accurately identify skin disease. The system is built to process skin lesion images and trained with machine learning algorithms to accurately diagnose skin diseases. Furthermore, the system is built to facilitate physicians to make data-based decisions for patients with the aim to reduce patient waiting time for dermoscopic screenings and receiving treatment (Almuayqil et al., 2022).

Medical image processing plays a crucial role in diagnosing skin disease. Libraries such as OpenCV, Scikit-Image, NumPy are used for processes such as segmenting disease lesions from normal skin areas and extracting features for instance lesion shape, color, and texture information. Feature extraction plays a significant role in image processing, it is used to extract characteristics from images as information for the subsequent classification task. Lastly, classification machine learning models such as k-Nearest Neighbour (KNN), Support Vector Machine (SVM) and Random Forest (RF) are employed to build a skin disease classification model (Aldera et al., 2022; Hatem, 2022). Other innovative techniques which employed convolutional neural network (CNN) models that were used for skin cancer classification and yield high accuracy of 95.18% (Shetty et al., 2022).

Skin diseases affect people of all skin tones, from fair skin to darker skin complexions. Hence, it is critical that machine learning algorithms are trained across diverse skin tones. However, many algorithms are trained with International Skin Imaging Collaboration (ISIC) dataset, that contains majority light skin tone samples but as consequence under-represent darker skin tone samples (Daneshjou et al., 2022). The use of imbalance dataset is a limitation of developed skin disease classifiers as the condition skin disease could present itself differently on darker skin and consequently lead to the inaccurate diagnosis (Kim et al., 2021). Moreover, another popular dataset used to train algorithms is HAM10000, it contains a large multi-source collection of 10015 dermoscopic images and for it, 53.3% of samples are pathologically verified of its disease (Tschandl et al., 2018). Additionally, HAM10000 dataset is imbalanced (Shetty et al., 2022). This dataset could potentially contain label noise since close to half of the samples do not have pathological confirmation. Label noise is a major concern as this could affect the reliability and accuracy of the algorithm.

II. Problem Statement

The existing skin disease classification models are trained with a dataset that has limited instances of dark skin tone patients. Consequently, the model developed is potentially biased and leads to inaccurate diagnosis for skin diseases especially patients with dark skin tone. Furthermore, skin disease manifests itself differently across different skin tones. Moreover, skin diseases appear visually similar thus is difficult to classify. Hence, there is a need to develop fair skin disease classification models for all people with different skin tones.

III. Research Questions

This research aims to investigate the following research questions:

1) What machine learning models can identify malignant and benign skin disease using skin lesion images across different skin tones ?
2) What machine learning algorithms can be used to classify different types of skin diseases?

IV. Research Objective

The objective of this study is as follows:

1) To develop machine learning models that are classify skin malignant and benign skin disease on different skin tones.

2) To build a multi-class classification model to identify four skin diseases namely, melanocytic nevi, seborrheic keratosis, verruca vulgaris, and basal cell carcinoma on different skin tones.

The building of the proposed model requires multiple image pre-processing techniques that includes resizing, colour transformation, normalization, segmentation, and feature extraction. Next, machine learning models such as RF, Decision Tree (DT), SVM, Naïve Bayes (NB), KNN, Extreme Gradient Boosting (XGBoost) algorithms are deployed to construct skin disease classification models and evaluated with evaluation metrics accuracy, precision, recall, F1-measure to determine the best performing model.

The main contribution of this proposed work is to ensure skin classification models are built to account for people of all skin tones. Such that models can be used universally, and could accurately predict skin disease on all skin tones. This paper is organized into the following sections. Section 2 discusses the previous work of skin diseases image processing and classification. Section 3 presents the methodology. Section 4 shows the obtained results and Section 5 discussion and conclusion of this research work.

V. Literature Review

There is multiple research that combines image processing techniques and machine learning algorithms to build skin disease classification models for accurate diagnosis. These models include binary and multi-class classification models. The following are the literature surveys conducted to explore factors like disease examined, dataset used, feature extracted and machine learning model performance achieved.

Hatem (2022) proposed a binary skin lesion classification model that implements KNN to differentiate normal skin and malignant skin lesions. The researcher's choice of using KNN is due to its time efficiency, good performance, and interpretability. The skin lesion images was segmented using adaptive thresholding. Features extraction methods implemented were mean (fast Fourier transform), standard deviation, histogram-based mean and standard deviation, edge-based pixel count of area and hole, edge-based logarithmic pixel count of area and hole to extract texture features. KNN algorithm was deployed to predict malignancy of a skin lesion. The model achieved 98% accuracy. This suggests that textural features and KNN are suitable for building skin classification algorithms.

Hegde et al. (2018) studied several machine learning methods to classify three common skin diseases chronic eczema, lichen planus and plaque psoriasis. The dataset consists of 310 images obtained from a dermatology clinic. Red, Green, Blue (RGB) color features and Gray Level Co-occurrence Matrix (GLCM) texture features are extracted from the image. Color, texture, and the combination of both features are used to train models. SVM and combined features result the accuracy of 81.61%. While discriminant Analysis (LDA) and color features achieved the same accuracy of 81.61%. This shows that color features are an important contribution to build accurate skin classification models. It also suggests that LDA as a promising skin classification algorithm.

Aldera et al. (2022) proposed a multi-class model to diagnose four different skin diseases namely, acne, cherry angioma, melanoma, psoriasis using image processing and machine learning techniques. A combination of two datasets are used in this study, which are Dermnet NZ dataset and Atlas dermatologico. The study used Otsu's thresholding technique to segment lesion regions from the skin. The feature extraction is performed using Gabor, Entropy for extract texture information and Sobel method to detect image's edge features from the images. This work used traditional machine learning algorithms to classify diseases which includes SVM, RF and K-NN. The proposed model result showed SVM to obtain the highest accuracy of 90.7%, precision of 91%, recall of 90.8% and F1-Score of 90.8% in comparison to RF and KNN.

Other than traditional machine learning, recent studies also utilize deep learning models to build skin classification model. Shetty et al. (2022) applied machine learning and CNN techniques to classify seven classes of skin lesion images. The dataset used in this study is HAM10000 dataset that consists of 10015 images. The images are pre-processed by resizing and augmented using horizontal flip augmentation method. The color, shape, texture properties of skin lesions are quantified using color histogram, Hu Moments, and Haralick Texture respectively. Based on the evaluation performance parameters, the proposed CNN model performed the best in terms of high accuracy of 95.18% among the others. However, given that the dataset is imbalanced, the suitable parameters are precision, recall and F1-scores. The random forest machine learning model performs better based on precision, recall and F1-scores with 94%, 94%, 94% in comparison to CNN model 88%, 85%, 86% respectively.

Salem Ghahfarrokhi et al. (2022) proposed a method for classifying skin lesions and diagnosis of melanoma using dermoscopic images. The dataset used for the The skin lesion image are segmented using Online Region-based Active Contour Model (ORACM). Given that the lesion shape is irregular and its borderlines are asymmetry, the author employed non-linear analysis consisting of generating time series, fractal dimension, lyapunov exponent, entropy to extract non-linear indices. Different texture and GLCM features were also extracted from the image. The research includes five machine learning models which are KNN, SVM, Fitting neural network (Fit net), Feed-Forward neural network (FF net) and Pattern network (Pat net) with combination of selected nonlinear indices, texture, GLCM texture for the classification task. The evaluation parameter is based on fivefold cross-validation. The Pat net models achieved the highest accuracy among the evaluated models with accuracy of 99.25%.

Based on the literature review, there are multiple methods, pre-processing, segmentation, feature extraction methods that are crucial to construct an accurate skin disease classification model. Additionally, there are a wide range of machine learning models studied for the skin classification task and able to perform with high accuracy. However, the dataset used in these study lacks the diversity of skin tone images and does not emphasize on building models suitable to diagnose skin disease for all skin tones.

VI. Motivation

Due to the limitation of existing algorithms used to diagnose different skin diseases, this research aims to build a model that diagnoses skin diseases such as malignant or benign skin diseases across diverse skin tones. The dataset used for this research is a public dataset known as the Diverse Dermatology Images (DDI) dataset. All instances in the dataset are expertly curated, skin tone was labelled based on physical examinations during in-person clinic visit by cross referencing against demographic photos and are its disease are pathologically confirmed samples. The dataset was curated to meet the limitation of other skin disease dataset. The DDI dataset has a fair distribution of different skin tone samples. Based on the Fitzpatrick skin type (FST) the dataset has a total 656 images, including FST I-II (light skin tone) 208 images, FST III-IV (medium olive skin tone) 241 images, and FST V-VI (dark skin tone) 207 images

VII. Methodology

This study demonstrates two types of classification model, one of it is a binary classification model, the other is a multi-class classification model. The binary classification model proposed classifies malignant and non-malignant skin disease. The multi-class classification model proposed to diagnose several skin diseases such as melanocytic nevi, seborrheic keratosis, verruca vulgaris, and basal cell carcinoma on different skin tones. The following points show details of the dataset and explanation of the image processing techniques used.

7.1 Dataset

The Diverse Dermatology Image (DDI) dataset consists of skin lesion images diagnosed in Sandford Clinics from 2010 to 2020. The skin tones are determined using chart review of the in-person visit and consensus review by two board-certified dermatologists, while the skin diseases are biopsy proven diagnoses (Daneshjou Roxana et al., 2022). The total images in the dataset is 656.

7.1.1 Dataset for Binary Classification

All 656 images are used to train the binary classification models. The dataset is breakdown is as follows:

Table 1: Dataset distribution of malignant or benign instances by each skin tone category.

Disease	FST I – II (light skin tone)	FST III – IV (medium olive skin tone)	FST V – VI (dark skin tone)	Total
Malignant	49	74	48	171
Benign	159	167	159	485
Total	208	241	207	656

7.1.2 Dataset for Multi-Class Classification

From the DDI dataset, a subset of four different disease classes namely melanocytic nevi, seborrheic keratosis, verruca vulgaris, and basal cell carcinoma was selected to build the multi-class classification models.

Table 2: Dataset distribution of four classes of skin disease by each skin tone category.

Disease	FST I – II (light skin tone)	FST III – IV (medium olive skin tone)	FST V – VI (dark skin tone)	Total
melanocytic nevi	47	49	23	119
seborrheic keratosis	21	18	19	58
verruca vulgaris	26	7	17	50
basal cell carcinoma	7	34	0	41
Total				268

7.2 Image pre-processing

Image pre-processing is an important step that manipulates and analyses input images. The main objective of image processing is to enhance the image's quality and extract features such that it is interpretable by the machine. The following describes the image processing techniques used in this work.

- **Resizing**- All images are resized to width and height of 40x40.
- **Median filter** – The median filter was applied. The median filter is a type of non-linear filter that is used to smooth an image by replacing the intensity of each pixel with the median intensity of a set of neighbouring pixels. It has the advantage of preserving the edges of the image (Aldera et al., 2022).
- **Colour conversion**: The images are converted to for the segmentation process and extraction process.
- **Normalization**: The pixel values of the image are normalized between 1 and 0. Normalizing the pixel values to a common range can improve the model's ability to learn from the data.

7.3 Image Segmentation

The segmentation task is a critical part to extract the lesion region of the skin. For this process, Otsu's thresholding was applied to the grayscale image. The choice of using Otsu's technique is due to its simplicity and effectiveness. The Otsu's algorithm compares the minimized weighted between-grey class variance to determine the optimal threshold value, subsequently the lesion can be segmented from the normal skin region.

7.4 Feature Extraction

Feature extraction is a technique used to extract features information from the image as explanatory variables for the classification model. In this work, we focused on texture features that describe the size, colour, brightness, slope, smoothness characteristics of the skin lesion image. The feature extraction was performed using Gabor technique, linear filter that extract texture features. The parameters used for Gabor filter is as follows:

Table 3: The parameter for Gabor feature extraction.

Feature Extraction Technique	Parameters	Value
Gabor	λ : wavelength of the cosine multiplier θ : frequency of alternations in degrees ϕ : Phase offset of the sinusoidal function σ:Sigma/standard deviation of the Gaussian envelop γ : Spatial aspect ratio and specifies the ellipticity of the support Gabor function	Pi/2.0 0, 45, 90, 135 Pi/4 0.5 0

7.5 Cross-validation

This study employed cross-validation to use all data for training and testing to validate the algorithm. The ‘k’ folds selected was five. The dataset is randomly partitioned into five groups, each group is taken as the test dataset while remaining groups are taken as the training dataset, this is repeated five times. The evaluation measures for five groups are averaged.

7.6 Classification

In this work, supervised machine learning algorithms are used to classify labelled skin diseases into distinct known class/classes. For proposed model used classifiers such as RF, DT, SVM, NB, KNN, XGBoost.

7.6.1 Binary Classification

The proposed binary model was developed to classify skin lesions into benign and malignant.

7.6.2 Multi-Class Classification

The multi-class classification model was developed to classify skin lesions into four classes of skin diseases, namely, melanocytic nevi, seborrheic keratosis, verruca vulgaris, and basal cell carcinoma.

Table 4: Classifier parameters

Classifier	Parameters Values
RF	Default
DT	Default
NB	Default
SVM	‘Kernel function’ is ‘RBF’
KNN	Number of nearest neighbours = 3
XGBoost	n_estimators = 110, max_depth=300, min_child_weight=1, njob=16

7.7 Evaluation metrics

The model performance was evaluated with accuracy, precision, recall and F1-score metrics for both binary and multi-class classification. Receiver Operator Characteristic – Area Under Curve (ROC-AUC) metric was used to evaluate the ability for the binary model to distinguish between the classes. The metric function are formulated as (1), (2), (3), and (4): Where TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

$$\text{Accuracy} = (TP+TN) / (TP+FP+TN+FN) * 100\% \quad (1)$$

$$\text{Precision} = TP / (TP+FP) \quad (2)$$

$$\text{Recall} = TP / (TP+FN) \quad (3)$$

$$\text{F1-score} = 2*(\text{precision}*\text{recall} / \text{precision} + \text{recall}) \quad (4)$$

VIII. Results

8.1 Binary Classification

Table 5: The obtained results of the evaluation metrics of the designed classifiers performing a 5-fold cross-validation binary classification model.

Classifiers	Accuracy	Precision	Recall	F1-Score	AUC-ROC
RF	74.23	0.58	0.04	0.14	0.58
DT	61.11	0.28	0.32	0.30	0.52
SVM	74.23	0.40	0.01	0.02	0.55
NB	62.03	0.30	0.34	0.32	0.57
KNN	72.71	0.30	0.06	0.10	0.53
XG Boost	73.01	0.40	0.09	0.15	0.54

8.2 Multi Class Classification

Table 6: The obtained results of the evaluation metrics of the designed classifiers performing a 5-fold cross-validation multi-class classification model.

Classifiers	Accuracy	Precision	Recall	F1-Score
RF	0.39	0.24	0.25	0.24
DT	0.35	0.23	0.29	0.26
SVM	0.44	0.11	0.25	0.15
NB	0.35	0.27	0.30	0.26
KNN	0.41	0.18	0.24	0.17
XG Boost	0.41	0.30	0.28	0.26

IX. Discussion

For the binary classification models, it is observed that RF has the best performance in classifying malignant and benign skin diseases over others in accuracy, precision and AUC-ROC of 74.23%, 0.58 and 0.58 respectively. Followed by SVM, XGBoost, KNN. However, for the multi-class skin disease classification models, it is observed that all machine learning models proposed performed poorly. Overall, both binary and multi-class models performed poorly in comparison to existing models. Here we discuss the likely reasons for such results.

From the pre-processing procedure, the poor result could likely be the insufficient data. To overcome the problem, data augmentation technique could be implemented to generate new “data” by Horizontal Flip augmentation. This allows models to learn more differentiating characteristic features in comparison to models without data augmentation (Shetty et al., 2022). Data augmentation increases sample size by rotating the images in horizontal direction. Moreover, other histopathologically proven skin lesion dataset could be used in combination with the DDI dataset used in this research to increase data size.

Next, the possible reason for low accuracy is insufficient input data. More features from images could be extracted, such as, color pigmentation information. The study conducted by Hegde Parameshwar R. et al. (2018) demonstrated that skin lesion image’s Red, Green, Blue (RGB) colour features are a major aspect for building skin disease classification models. Additionally, Gray Level Co-occurrence Matrix (GLCM) is a texture feature extraction used in combination with colour features as input features for training classifiers. The performance of the classifiers achieved high accuracy (Hegde Parameshwar R. et al., 2018). Further studies between Gabor and GLCM texture feature extraction performance could be conducted to determine the most suitable methods.

Lastly, research showed using a combination of algorithm such as SVM and RF algorithm together achieved a higher accuracy compared when it is used individually (Murugan et al., 2021). The study demonstrated that when RF and SVM were individually used for skin disease classification with GLCM texture features, the model accuracy is 76.36% and 87.8% respectively. However, when used in combination, the accuracy improved with 89.31%.

X. Conclusion

The most critical step in medical health care is the proper diagnosis of disease therefore it is essential to build high accuracy models for skin disease classification. Moreover, skin algorithms developed must be capable of diagnosing skin disease on different skin tones. Therefore, using the suitable methodology for image pre-processing, feature extraction and using the right machine learning models is needed to build accurate skin disease classifiers. Although this study did not manage to develop high accuracy models, this study contributes by raising awareness in building fair skin disease classification models.

XI. References

Aldera, S. A., Tahar, M., & Othman, B. (2022). A Model for Classification and Diagnosis of Skin Disease using Machine Learning and Image Processing Techniques. *IJACSA) International Journal of Advanced Computer Science and Applications*, *13*(5). www.ijacsa.thesai.org

Daneshjou Roxana, Vodrahalli Kailas, Novoa Roberto A, Jenkins Melissa, Liang Weixin, Rotemberg Veronica, Ko Justin, Swetter Susan M, Bailey Elizabeth E, Gevaert Olivier, Mukherjee Pritam, Phuang Michelle, Yekrang Kiana, Fong Bradley, Sahasrabudhe Rachna, Allerup Johan A.C, Okata-Karigane Utako, Zou James, & Chiou Albert S. (2022). Disparities in dermatology AI performance on a diverse, curated clinical image set. *Science Advances* , *8*.

Hatem, M. Q. (2022). Skin lesion classification system using a K-nearest neighbor algorithm. *Visual Computing for Industry, Biomedicine, and Art*, *5*(1). https://doi.org/10.1186/s42492-022-00103-6

Hegde Parameshwar R., Shenoy Manjunath M., & Shekar B.H. (2018). Comparison of Machine Learning Algorithms forSkin Disease Classification Using Color andTexture Features. *INTERNATIONAL CONFERENCE ON ADVANCES IN COMPUTING, COMMUNICATIONS AND INFORMATICS (ICACCI)*, 1825–1828.

Murugan, A., Nair, S. A. H., Preethi, A. A. P., & Kumar, K. P. S. (2021). Diagnosis of skin cancer using machine learning techniques. *Microprocessors and Microsystems*, *81*. https://doi.org/10.1016/j.micpro.2020.103727

Salem Ghahfarrokhi, S., Khodadadi, H., Ghadiri, H., & Fattahi, F. (2022). Malignant melanoma diagnosis applying a machine learning method based on the combination of nonlinear and texture features. *Biomedical Signal Processing and Control*, *80*(2023). https://doi.org/10.1016/j.bspc.2022.104300

Shetty, B., Fernandes, R., Rodrigues, A. P., Chengoden, R., Bhattacharya, S., & Lakshmanna, K. (2022). Skin lesion classification of dermoscopic images using machine learning and convolutional neural network. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-22644-9