Malignant skin melanoma detection using image augmentation by oversampling in nonlinear lower-dimensional embedding manifold

Olusola Oluwakemi ABAYOMI-ALLI[1] ![ORCID], **Robertas DAMAŠEVIČIUS**[1,*] ![ORCID],
Sanjay MISRA[2] ![ORCID], **Rytis MASKELIŪNAS**[3] ![ORCID], **Adebayo ABAYOMI-ALLI**[4] ![ORCID]
[1]Department of Software Engineering, Kaunas University of Technology, Kaunas, Lithuania
[2]Department of Computer Science and Communication, Ostfold University College, Halden, Norway
[3]Faculty of Applied Mathematics, Silesian University of Technology, Gliwice, Poland
[4]Department of Computer Science, Federal University of Agriculture Abeokuta, Nigeria

Abstract: The continuous rise in skin cancer cases, especially in malignant melanoma, has resulted in a high mortality rate of the affected patients due to late detection. Some challenges affecting the success of skin cancer detection include small datasets or data scarcity problem, noisy data, imbalanced data, inconsistency in image sizes and resolutions, unavailability of data, reliability of labeled data (ground truth), and imbalance of skin cancer datasets. This study presents a novel data augmentation technique based on covariant Synthetic Minority Oversampling Technique (SMOTE) to address the data scarcity and class imbalance problem. We propose an improved data augmentation model for effective detection of melanoma skin cancer. Our method is based on data oversampling in a nonlinear lower-dimensional embedding manifold for creating synthetic melanoma images. The proposed data augmentation technique is used to generate a new skin melanoma dataset using dermoscopic images from the publicly available PH^2 dataset. The augmented images were used to train the SqueezeNet deep learning model. The experimental results in binary classification scenario show a significant improvement in detection of melanoma with respect to accuracy (92.18%), sensitivity (80.77%), specificity (95.1%), and F1-score (80.84%). We also improved the multiclass classification results in melanoma detection to 89.2% (sensitivity), 96.2% (specificity) for atypical nevus detection, 65.4% (sensitivity), 72.2% (specificity), and for common nevus detection 66% (sensitivity), 77.2% (specificity). The proposed classification framework outperforms some of the state-of-the-art methods in detecting skin melanoma.

Key words: Malignant melanoma, skin cancer recognition, data scarcity, data augmentation, oversampling, transfer learning, deep learning

1. Introduction

Malignant melanoma is a deadly form of skin cancer that is responsible for over 20,000 deaths in Europe with close to 100,000 new cases recorded each year [1]. World Health Organization (WHO) claimed that approximately 13% of the global mortality rate is associated with cancer-related diseases and this is still on the rise as future estimate predict that 12 million more are at risk of death by the end of 2030 [2]. Recent studies have shown that early detection of this disease is the most critical prognostic factor for survival [3]. Presently, the number of cases examined by hospital pathologists rely on the visual examinations. Sensitivity of dermoscopic melanoma recognition in most cases is less than 80% in routine clinical settings [4]. Differentiating

---

*Correspondence: robertas.damasevicius@ktu.lt

between melanoma and nonmelanoma lesions is very hard by a nonexpert pathologist. The manual method of detecting skin lesions is human-labor intensive, prone to errors considering the level of expertise and time for visual exploration required, etc. [5]. Therefore there is a need to improve the computer-aided methods to effectively diagnose skin melanoma cancer as early as possible thereby increasing the survival rate [6].

Feature extraction from dermoscopy images using handcrafted texture and color features could be very challenging especially when comparing melanomas from nevi due to visual similarity, intraclass variations, and artifacts [7]. As a result, the analysis of dermoscopic images in real-time continues to remain an attractive research area for artificial intelligence (AI) researchers. Current research endeavors focus on the development of the AI methods that would serve doctors and patients through intelligent detection and diagnosis of skin cancer diseases [8]. Deep learning (DL) methods have played a vital role in skin disease detection and classification [9]. Several DL architectures were proposed in previous work for detecting skin melanoma such as residual network ResNet50 [4], attention residual learning [10], AlexNet [11], VGG [12], MobileNet, InceptionV3, etc. [5].

The application of advanced AI methods usually requires a large dataset, which is one of the main challenges in developing clinical level AI systems [13]. Apart from the shortage of publicly available datasets used for skin disease classification, the datasets also suffer from class imbalance, noise, high dimensionality of features, and inconsistency of image sizes and resolution. Research attention has been targeted towards developing larger image datasets, but there are challenges such as lack of metadata and imaging standards in dermatology, privacy issues associated with patient's images, image copyright, diagnostic ground-truth, etc. In many real-world cases, big datasets are not available, and when the dataset is too small, a neural network is trained ineffectively [14] thus leading to suboptimal and overfit solutions.

Data augmentation can help to improve accuracy of learning algorithms [15] in case of data scarcity. Some of the most commonly used data augmentation techniques in skin melanoma include random image transformation; horizontal and vertical flips [16], random contrast and random brightness, cropping [17], rotation, zoom in and out, histogram scaling, etc. [18]. The extraction of relevant features from these small and imbalanced data requires purposeful applications of thorough data analysis techniques. Methods such as feature space sampling have been used to solve the problem of imbalanced and small datasets and include random undersampling, clustering undersampling, random oversampling, SMOTE (Synthetic Minority Oversampling Technique) [19], etc.

The application of data augmentation techniques has played a positive role in improving the performance of the classifiers through leveraging on small data samples and generating synthetic samples such as fixed and random rotation angle approaches [11], color space transformation from RGB (red, green, blue) color space to HSV (hue, saturation, value) color space [20] to increase the number of skin images available for training. Therefore, data augmentation can be described as the act of enriching training data samples and variance of a given dataset for better generalization and capturing the underlying distribution of the training data [21, 22] .

The novelty and main contribution of this paper is an enhanced data augmentation technique based on the covariant SMOTE [23] technique to generate synthetic training data, improve the performance of the learning model, and reduce overfitting of a classifier for overcoming the challenges of limited dermatological images and class imbalance.

The remaining parts of the paper are as follows. Section 2 describes in detail the methods and material of our proposed model. Section 3 presents the implementation and experimental results obtained with detailed discussion. The paper concludes in Section 4 with highlights on contributions.

2. Methods and materials

2.1. Outline of the proposed model

The paper presents a skin image classification framework based on image augmentation using image embedding into lower dimensionality manifold and performing a variant of SMOTE for oversampling, and image classification using SqueezeNet [24]. The framework is divided into three stages as shown in Figure 1 and explained in more detail in the following subsections.

2.2. Preprocessing stage

Data preprocessing is very important for ensuring data cleaning and preparation. The need to avoid inaccuracies of classifier training requires that the convolutional network is able to differentiate between low contrast among skin lesion and nearby healthy skin, uneven border, and skin artifacts such as hairs, black frames, etc. [25].
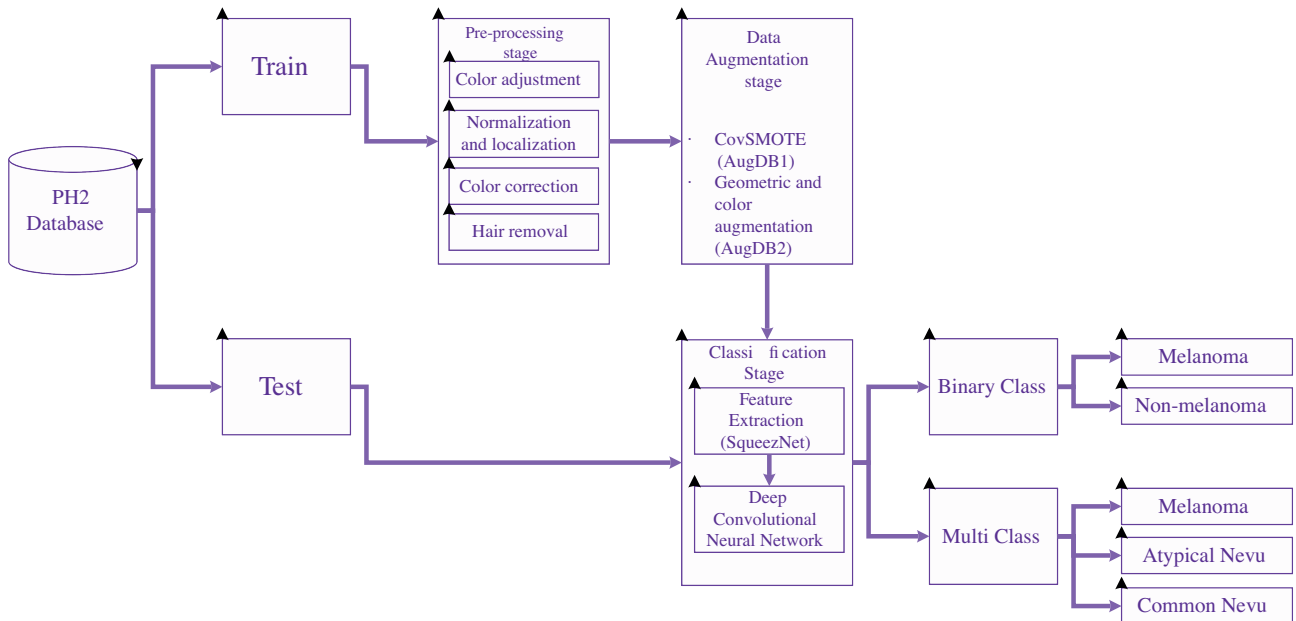

Figure 1. Flow diagram of the proposed method.

Before classification, we perform the following stages.

1. First, we adopt the hybrid color space to combine different color components to enhance the color component efficiency to reduce the measure of correlation dependency between color components. A conversion of images RGB color space to the CIE 1976 was done using standard illuminant D65, [0.9504, 1.0000, 1.0888], which simulates noon daylight with a color temperature of 6504 K. The L*a*b* (Lab) colors space was conducted using the Hering's theory for its effectiveness in measuring small color changes [26]. The theory shows that only red (R), green (G), blue (B), and yellow (Y) can characterize the hue component [27]. The LAB color space is a 3-dimensional color system with dimension L* for lightness, and a* and b* for color, where L* is the luminance presented in Eq. (1), where value 0 is represented as perfect black, and value 100 is assigned to perfect white. The a* and b* axes represent redness/greenness

and yellowness/blueness as expressed in Eq. (1), and lightness of a pixel is expressed in Eq. (2–4):

$$L^* = \begin{cases} 116\left(\frac{Y}{Y_W}\right)^{\frac{1}{3}} - 16 & \text{for } \frac{Y}{Y_W} > 0.008856 \\ 903.3\left(\frac{Y}{Y_W}\right) & \text{for } \frac{Y}{Y_W} \leq 0.008856 \end{cases} \tag{1}$$

$$a^* = 500\left[f\left(\frac{X}{X_W}\right) - f\left(\frac{Y}{Y_W}\right)\right] \quad b^* = 200\left[f\left(\frac{Y}{Y_W}\right) - f\left(\frac{Z}{Z_W}\right)\right] \tag{2}$$

$$f(\varsigma) = \begin{cases} (\varsigma)^{\frac{1}{3}} & \text{for } \varsigma \geq 0.008856 \\ 7.787(\varsigma) + \frac{16}{116} & \text{for } \varsigma \geq 0.008856 \end{cases} \tag{3}$$

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412 & 0.357 & 0.180 \\ 0.212 & 0.715 & 0.072 \\ 0.019 & 0.119 & 0.950 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \tag{4}$$

2. Next, to minimize illumination heterogeneity and also to increase the edges of the skin lesions, we implemented a high-pass filter. The high-pass filter is derived by removing the values of the weighted Gaussian blurred image from the original image as expressed in Eq. (5).

$$I_f = I - {}_wG_\sigma(I), \tag{5}$$

where I is input image, I_f is high-pass filtered image, G_σ is two-dimensional Gaussian kernel, σ is Gaussian window width, and $w \subset [0,1]$ is weighting factor that controls the degree of enhancement.

3. We applied the contrast limited adaptive histogram equalization (CLAHE) algorithm to perform image contrast enhancement. The CLAHE algorithm optimizes the contrast enhancement on local image data thus effectively reducing the global noise. Furthermore, standard histogram equalization is applied to each of the nonoverlapping contextual regions. Therefore, each region is contrast-enhanced locally which is followed by clipping and median filtering. CLAHE algorithm can prevent gray-level peaks by examining regional and global histogram equalization [28] . To determine two parameters of the CLAHE, we use an approach proposed in [29] and an example of the results of image enhancement is presented in Figure 2.

4. Hair is detected with Frangi filter [30]. The Hessian matrix H of the image I is derived from the 2nd order Gaussian derivative of I at point (x,y) and scale σ as in Eq. (6):

$$H = \frac{d^2 I_\sigma}{dx^2} = I(x,y)\frac{d^2 G(\sigma, x, y)}{dxdy}. \tag{6}$$

The Frangi function is expressed in Eq. (7) as:

$$F(\sigma, x, y) = e^{\left(-\frac{R_B^2}{2\beta^2}\right)}\left(1 - e^{\left(-\frac{S^2}{2c^2}\right)}\right). \tag{7}$$

Figure 2. Example of image enhancement: image from PH^2 dataset (left); enhanced image (right).

We applied Frangi filter to segment all hairs from the background, while β is set to 0.5 and c is set to 15. Scale range is set between 0.01 and 7.01, increasing by 0.5 for each scale. We used inpainting based on the minimization Mumford–Shah functional [31] to fill in the image pixels removed by the mask as follows:

$$\min_{u,k} \left\{ \int_{\Omega} |u - f|^2 dx + \alpha \int_{\Omega \setminus K} |\nabla u|^2 dx + \lambda |K| \right\}, \tag{8}$$

where $\omega \rightarrow R^2$ is image domain; $f : \omega \rightarrow R^2$ is initial image function; $u : \omega \rightarrow R^2$ is unknown approximation, which is smooth everywhere in ω but can have a jump at subregions' edges K; $|.|$ is the Euclidean norm; $\alpha > 0$ and $\lambda > 0$ are parameters, which control smoothness and length of edges.

2.3. Image mapping to lower dimensional manifold space

The mapping of skin images to lower dimensional manifold space is performed as follows. Every pixel of the skin image is used to construct a high-dimensional manifold space. We adopted a differential manifold for constructing skin image manifold where its local space follows the Euclidean-style space as expressed in Figure 3.

For dimensionality reduction of the manifold space, we applied Gaussian distribution to generate low-dimensional space by describing distribution characteristics (T) of data points set within the high-dimensional manifold space. We minimized the contrast between distributions to finalize the operation of skin classification. We first applied random projection on a plane of image data point set using a discrete distribution.

Euclidean distance in Eq.(12) is used to determine the interval between points in a 2D plane projection of skin image data points and surrounding neighboring points with space L_p of x_i and x_j. K-Nearest Neighbor (KNN) map of the skin image pixels can be generated by the previously described steps, and the obtained KNN map represents the local features of the manifold data pointset. The skin image data point x_i corresponds to the center of the Gaussian distribution. Point x_i uses probability $P_{(j|i)}$ to select point x_j as its nearest neighbor. Probability $P_{(j|i)}$ is inversely proportional to the distance of x_j and can be described by Eq. (10):

$$L_2(x_i, x_j) = \left(\sum_{l=1}^{n} \left| x_i^{(l)} - x_j^{(l)} \right|^2 \right)^{\frac{1}{2}} \tag{9}$$

Start:

Euclidean space: $R^d = \{ Y \in X\}$

For ∀ point **in** X = ({ $X_1, X_2, ..., X_p$}, Y)

X ← d-dimensional topological manifold

V ← C^k differential structure

$V = \{(U_\alpha, \varphi_\alpha)\}_{\alpha \in I}$

The manifold M formed by {$U_\alpha : \alpha \in I$} is an open cover, and the mapping of $\varphi_\alpha : U_\alpha$ to

If ($U_\alpha \cap U_\beta \neq \emptyset$, && double mapping $\varphi_\alpha \circ \varphi_\beta^{-1} : \varphi_\beta(U_\alpha \cap U_\beta)$ to $\varphi_\alpha(U_\alpha \cap U_\beta)$)

Then $(U_\alpha, \varphi_\alpha)$ is compatible with (U_β, φ_β)

End

Figure 3. Image mapping to manifold space.

$$P_{j|i} = \frac{exp(-\|x_i - x_j\|^2/2\sigma_i^2)}{\sum_{k \neq i} exp(-\|x_i - x_k\|^2/2\sigma_i^2)}, \tag{10}$$

where γ_i is the variance of the Gaussian distribution with x_i as the center-point. To prevent a crowding issue, we applied T-distribution to illustrate the local connection of the image points in the low-dimensional manifold. Suppose that the probability distribution of the neighborhood image data point set in the high-dimensional space can be characterized by the normal distribution $N(\mu, \sigma^2)$. The mean value of the T-distribution in the low-dimensional space can be obtained using expression in Eq. (11), while the variance of image points can be derived from Eq. (12) and T-distributed random variable is constructed using Eq. (13):

$$\overline{u} = \frac{1}{n} \sum_{i=1}^{n} u_i \tag{11}$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (u_i - \overline{u})^2 \tag{12}$$

$$t = \frac{\overline{u} - \mu}{\frac{s}{\sqrt{n}}}, \tag{13}$$

where u_i is the skin image points in the KNN domain, and $\overline{u}$ is their mean. A uniform symmetrical distance function Eqs. 14 and 15 sustains the symmetry of the two probability distributions in the 2-dimensional space:

$$q_{i|j} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq l}(1 + \|y_k - y_l\|^2)^{-1}} \tag{14}$$

$$p_{ij} = \frac{p_{j|i} + P_{i|j}}{2n}, \tag{15}$$

where $q_{(}i|j)$ is the probability distribution of the skin image points in the 2D-space.

We calculate the loss function using the Kullback–Leibler (KL) divergence with the aim of reducing the intervals between distribution of high-dimensional space $P_{(}j|i)$ and low-dimensional space $q_{(}i|j)$. The whole process for reducing dimensionality for the least value of the loss function is expressed in Eq. 19. In order, to manage the problem of overfitting resulting from limited data samples, a degree of confusion is also applied as expressed mathematically in Eqs. 16 and 17. The Shannon entropy $H(P_i)$ with p_i is defined in Eq. 18.

$$C = \sum_i \sum_j p_{j|i} \log \frac{p_{j|i}}{q_{i|j}} \tag{16}$$

$$Perp(p_i) = 2^{H(P_i)} \tag{17}$$

$$H(P_i) = -\sum_j p_{j|i} \log_2 p_{j|i} \tag{18}$$

The adjustment to the degree of confusion is proportional to the entropy thus, the value of entropy varies by adjusting the degree of confusion. We further construct the objective function by clustering the 2D space of both positive and negative samples of the skin image data points. Weight between points is set by Eq. (19).

$$P(e_{ij} = 1) = f(\|y_i - y_j\|^2), \tag{19}$$

where y_i and y_j is the two points in the low-dimensional space. The probability of a binary edge between two points is represented as $P(e_i j = 1)$. However, $P(e_i j = 1)$ becomes larger as the interval space between y_i and y_j gets closer and the mathematical expression with weight $w_i j$ is depicted in Eq. 20. The KNN diagram was further applied to obtain the positive sample set E and the negative sample set $\bar{E}$.

$$P(e_{i,j} = w_{i,j}) = p(e_{i,j} = 1)^{w_{i,j}} \tag{20}$$

For optimization, we maximize the probability of positive samples weighted edges in KNN graphs and reducing probability of negative samples in weighted edges of KNN graphs. Eq. 21 is transformed into Eq. 22:

$$O = \Pi_{(i,j)\in E}\, p(e_{ij} = 1)^{w_{ij}} \Pi_{(i,j\in\overline{E})}(1 - p(e_{ij} = 1))^{\gamma} \tag{21}$$

$$O = \sum_{(i,j)\in E} w_{i,j}(p(e_{i,j} = 1) \sum_{(i,j)\in E} \gamma(1 - p(e_{i,j} = 1)), \tag{22}$$

where γ is the weight of the negative samples. The computational complexity can be increased by a negative sample $\bar{E}$ thus becoming difficult to directly apply gradient descent for training. Consequently, we carefully choose a negative sampling algorithm, and using a randomly selected number of skin image points, we developed a negative sample with the noise distribution $P_n(j)$. Eq. 23 illustrates the objective function:

$$O = \sum_{(i,j)\in E} w_{i,j}(p(e_{i,j} = 1) + \sum_{k=1}^{M} \sum_{jk} \sim p_n(j)\gamma \log(1 - p(e_{i,j} = 1))). \tag{23}$$

2.4. Data augmentation stage

To perform data augmentation, we have adopted a modification of the Synthetic Minority Oversampling Technique (SMOTE) [19] method. SMOTE creates artificial instances by oversampling the minority class. SMOTE recognizes k-minority class neighbors that are near to each minority class sample by randomly selecting a point between the original sample and its neighbors as the artificial sample. We adopted a variant of SMOTE, called $SMOTE - cov$ [23], which uses the covariance matrix (CM) to detect dependency connection between attributes. From the estimated CM, we generate surrogate instances to achieve the balance between the minority class and the majority class. This process is discontinued once an equilibrium between the classes is achieved. The example results for PH^2 dataset in the manifold space are presented in Figure 4.
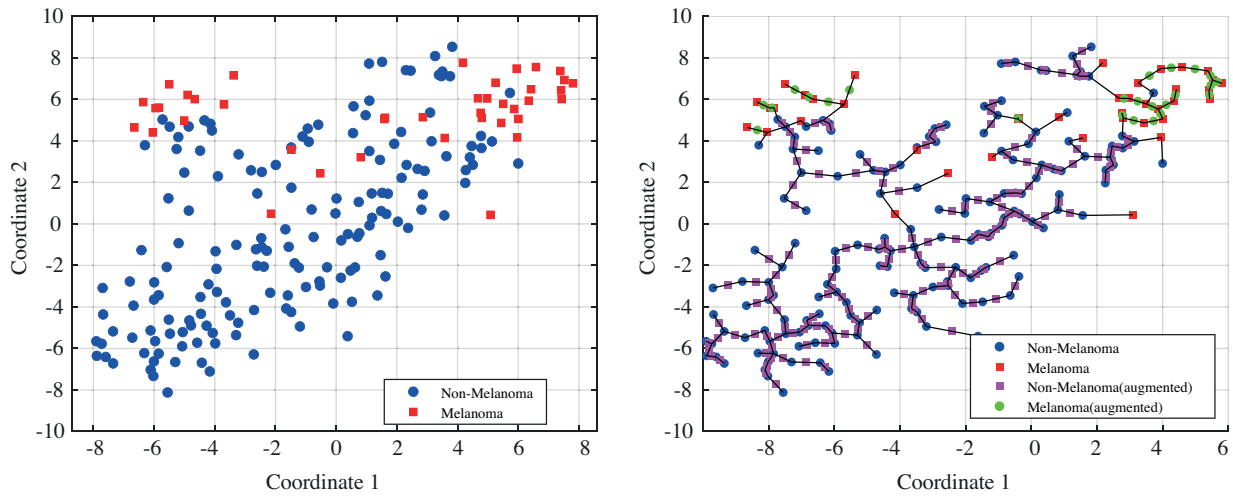

Figure 4. (a) Data class visualization of the classes in the PH^2 dataset; (b) Example of new instances generated in the manifold space using the PH^2 dataset. Green dots show the newly generated instances.

We use the same approach to generate new minority class instances in the lower-dimensional manifold space as described in the previous subsection. After new instances are generated, we transform them from the manifold space to the image space. Since the manifold space is not linear, we adopt the polynomial regression approach to find the appropriate transform in the vicinity of each new instance. To begin, we find the nearest n images in the manifold space M that are closest to the new instance X_i. To find a mapping from the manifold space to the color space $M : M \rightarrow R^3$, we perform a cubic polynomial regression as follows. Given n data points $(X_1, X_2, , X_n), X_i \rightarrow R^2$, we use the least-squares optimization method to regress the data to an m-th order polynomial as expressed in Eq. 24.

$$I = A_0 + A_1X + A_2X^2 + ... + A_mX^m, m < n \tag{24}$$

Here A_m are the coefficient matrices and the residual at each image I_i is given by: The above are solved to obtain a mapping $M = (A_0, A_1, , A_m)$. The sample generated images are illustrated in Figure 5.
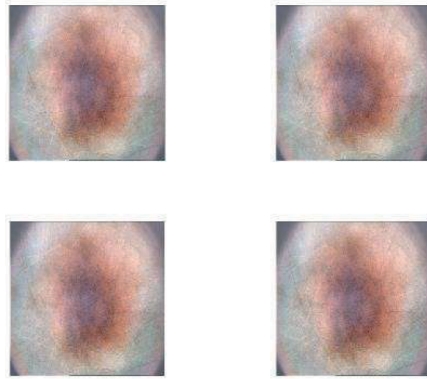
Figure 5. Examples of synthetic images generated by the proposed image augmentation technique.

2.5. Deep learning model

Deep learning methods have an important role in the automatic extraction of the relevant set of features and also successfully improving classification performance [32]. An effective deep learning method is critical to recognize dermoscopic feature patterns in the skin lesion area, such as streaks, globules, blotches, blue-white veils and pigment networks. For this study, we adopted the feature extraction method based on the SqueezeNet deep learning architecture as presented in Figure 6. We selected SqueezeNet, because it allows to achieve AlexNet level performance with with 50x fewer parameters [24].
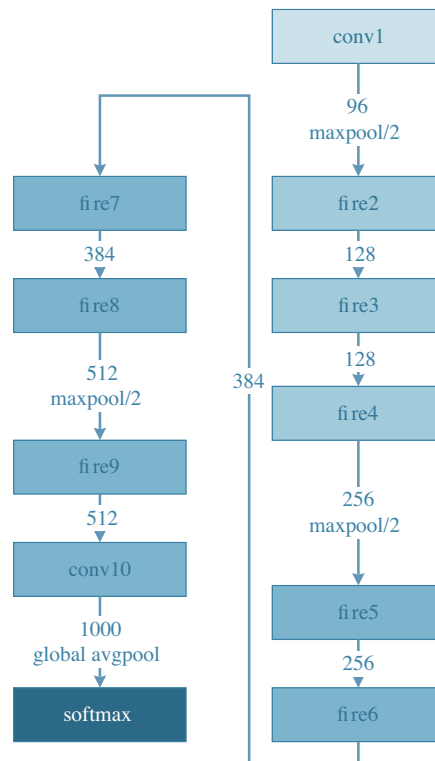

Figure 6. Architecture of SqueezeNet network.

SqueezeNet stacks multiple fire unit and a few pooling layers. The fire unit has the squeeze layer and the expand layer to keep the same feature map size. The squeeze layer uses 1×1 convolution kernels to reduce the parameters of network. The expand layer has a mix of 1×1 and 3×3 convolution filters. The results is a deep network with high performance and low number of parameters.

2.6. Performance metrics

Due to the problem of class imbalance associated with skin melanoma dataset, it is important to consider other performance measures apart from accuracy to optimize the performance measures. We evaluated the performance of the proposed method using other optimization metrics such as sensitivity (S_e), specificity (S_p), precision (P_r), and F-score.

3. Experimental results and discussion

3.1. Dataset

We conducted our experiment using PH^2 dataset [33] collected by the Dermatology Service of Hospital Pedro Hispano, Portugal. A total of 200 images are comprising of 3 classes of clinical diagnosis, namely common nevus, atypical nevus, and melanoma with a sample ratio of 80:80:40 samples, respectively. For the experiment, the PH^2 dataset was split into a training dataset (60%) and validation datasets (40%) were used for testing the trained network. From the training image samples, we generated two different databases using two data augmentation methods. The first data augmentation method was based on the covariant SMOTE technique which was used to generate a new database (AugDB-1) and also balancing the minority class (melanoma). We further used the conventional geometric data augmentation methods based on rotation, horizontal and vertical flipping, translation, etc. to generate the second database (AugDB-2). The summary of the experimental data used in this study is presented in Tables 1 and 2. The analysis and the performance results of our proposed model on these two data categories which are binary and multiclass classification.

Table 1. Summary of data categories for binary classification.

Data	Total images	Training set		Testing set	
		Melanoma	Nonmelanoma	Melanoma	Nonmelanoma
Original	200	24	96	16	64
AugDB-1	224	48	96	16	64
AugDB-2	472	236	236	16	64

Table 2. Summary of data categories for multiclass classification.

Data	Total images	Training set			Testing set		
		Melanoma	Atypical nevus	Common nevus	Melanoma	Atypical nevus	Common nevus
Original	200	24	48	48	16	32	32
AugDB-1	224	48	48	48	16	32	32
AugDB-2	472	236	236	236	16	32	32

3.2. Classification results

The implementation of this study was carried out on a workstation with Intel Core i5-5300U CPU @ 2.30GHz with 8GB RAM and 64-bit Windows 10 operating system. The tool used for implementing the proposed model was on MATLAB (MathWorks Inc, USA) Deep Learning Toolbox. The network was trained using the stochastic gradient descent (SGD) optimizer for 10 epochs with the best learning rate chosen from $10^{-4}, 10^{-5}, 10^{-6}$, piecewise learning rate schedule, the learning rate drop factor of 0.2, and the learning rate drop period of 2. The classification results for binary and multiclass are fully discussed in details below.

3.2.1. Classification results for binary class skin melanoma detection

The first test was conducted on the original dataset (PH^2) using the ratio of $60 : 40$ for training and testing. For the binary classification analysis, we categorized the original dataset into two: melanoma and nonmelanoma images. We train on 120 images which comprise 24 melanoma and 96 nonmelanoma images and tested on 80 test images (16 melanoma and 64 nonmelanoma) images as summarized in Table 1. The performance of our proposed model on the PH^2 dataset using the ImageNet pretrained network achieving the best accuracy of 0.925, the sensitivity of 0.875, specificity of 0.9375, and F1-score of 0.8235. The experiment was repeated 10 times and the mean performance metrics obtained were 0.8964 for accuracy, the sensitivity of 0.726, specificity of 0.9406, and F1-score of 0.7365.

The second experiment was based on using our proposed data augmentation method for generating a purely synthetic dataset for the minority class (melanoma). This newly created dataset (AugDB-1) was developed using 60% of the melanoma images (24 images) to generate additional 24 new synthetic images which finally summed up to a total of 144 training images consisting of 48 images for the melanoma and 96 images for nonmelanoma. These new training images were used to train the model and the remaining 40% were used for testing the model. The test results for the AugDB-1 dataset effectively improves the performance of classification with the best accuracy of 0.953, sensitivity at 0.8462, specificity at 0.9804, and F1-score at 0.88. The performance after repeated experiments gave a mean accuracy rate of 92.18%, sensitivity rate of 80.77%, specificity rate of 95.1%, and F1-score of 80.84% respectively.

To validate the effectiveness of our proposed data augmentation method, we applied the geometric and color transformation techniques which include horizontal flipping, scaling, rotation, translation, contrast adjustment, sharpened images, etc. Based on these techniques, a new training dataset was generated (AugDB-2) with a total of 472 images consisting of 236 melanoma and 236 nonmelanoma images. The experiment results obtained showed that pretraining the model using the AugDB-2 gave an accuracy below 90% due to the problem of overfitting. The performance comparison results for the three databases are presented in Table 3. The confusion matrix of the best classification results for the multiclass classification scenario is given in Figure 7a.

Table 3. Performance comparison for binary classification scenario.

Data	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1-score (%)
Original	89.64	76.12	72.58	94.06	73.65
AugDB-1	92.18	81.13	80.77	95.1	80.84
AugDB-2	89.26	82.87	58.75	96.88	67.83

3.2.2. Classification results for multiclass skin melanoma detection

We further carried out a multiclass detection for the original three classes of the PH^2 dataset. The summary of experimental results for three classes: melanoma, atypical nevus, and common nevus is presented in Table 4.

Table 4. Performance comparison for multiclass classification scenario.

Dataset	Acc (%)	Melanoma				Atypical nevus				Common nevus			
		Se (%)	Sp (%)	Pr (%)	F1 (%)	Se (%)	Sp (%)	Pr (%)	F1 (%)	Se (%)	Sp (%)	Pr (%)	F1 (%)
Original	60.4	64.2	93.6	73.8	67.4	58.8	68.9	68.9	53.8	63.9	73.4	62.0	62.5
AugDB-1	66.4	89.2	96.2	84.1	74.9	65.4	72.2	61.3	62.5	66.0	77.2	66.6	65.4
AugDB-2	62.0	61.5	73.0	76.1	57.6	57.4	61.9	54.0	58.4	56.1	75.2	64.0	58.3

From Table 4, we can see the improvement in the performance of our model trained on AugDB-1 for melanoma detection with respect to the accuracy, sensitivity, specificity, and F1-score as 66.4%, 89.2%, 96.2%, and 74.9%, respectively. While the detection rate for atypical nevu shows a sensitivity rate of 65.4%, specificity rate as 72.2%, and F1-score as 62.5%; For common nevu detection rate gave a sensitivity as 66.0%, specificity as 77.2%, and F1-score of 65.4%. The confusion matrix of the best classification results for the multiclass classification scenario is given in Figure 7b. Our proposed classification framework was able to effectively detect melanoma images as the sensitivity rate shows an impressing improvement when compared to the other classes.
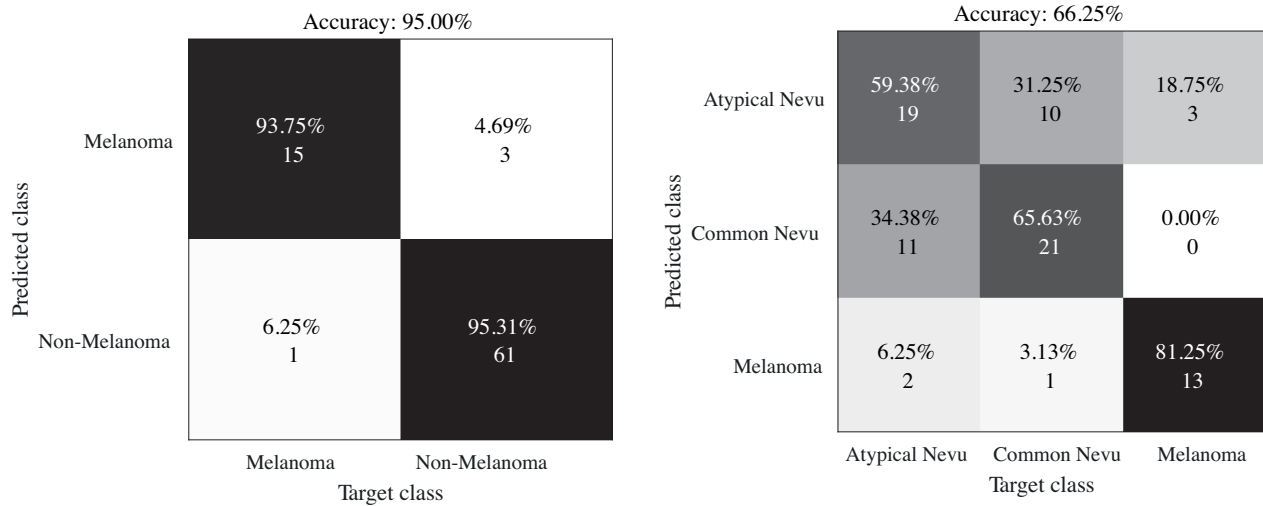

Figure 7. Confusion matrices of binary (a) and multiclass (b) classification results.

3.2.3. Comparison with related work

To validate the performance of our proposed model, we compared our work with the experimental outcome from previously published work using PH^2 dataset as shown in Table 5. In some of the cases our proposed approach improves significantly over some previous results. We also can also emphasize that the computational requirements of our approach were minimal with the time complexity for both binary and multiclass.

Our proposed method delivered an improved performance over some of the results identified in the previous studies (see Table 5). However, the limitation of the proposed method is that sensitivity and specificity

rate still requires improvement as we intend to further explore other deep learning architectures to improve the overall skin detection rate.

Table 5. Comparison with the state-of-the-art works using PH^2 dataset.

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	Reference
Combined Gaussian method with color space transformation. Classification is done using deep CNN based on Inception V3.	97.74	97.39	98.10	Khan et al. [34]
Feature similarity measurement (FSM) codebook learning algorithm on SVM classifier	91.9	92.5	91.3	Hu et al. [35]
SVM classifier with the linear kernel	86.07	78.93	93.25	Adjed [36]
Fuzzy C-means and Markov random field methods	94.0	93.2	98.0	Eltayef et al. [37]
Multiscale lesion-biased representation and joint reverse classification	92.0	87.5	93.13	Bi et al. [38]
Deep CNN-based using SqueezeNet and covariant SMOTE augmentation in low dimensional manifold	92.18	80.77	95.1	Proposed method

4. Conclusion

Melanoma accounts for the most skin cancer-related deaths, and current research is targeted at early detection of this disease. The growth of research endeavors in this area has experienced some constraints, which include a small available dataset for skin cancer detection, class imbalance of skin tumors, limited labeled data, poor standardization of clinical images, etc., which have resulted in the poor performance of the classifiers. Based on these challenges, this study improved the classification performance of melanoma skin disease by carefully applying the proposed model. The proposed model is an effective data augmentation method that is based on covariant SMOTE to solve the class imbalance problem. The effectiveness of the proposed data augmentation method has been illustrated through comparison with conventional data augmentation techniques and other existing approaches.

The obtained results for multiclass classification scenario demonstrate that the developed neural network architecture can be used to assist dermatologists during a melanoma screening process. The effectiveness and reliability of the proposed method are proven through comparisons with existing techniques. The proposed method shows the promising result and can successfully be integrated into a low-resource constraint device and adequately solve the real-time detection of skin melanoma detection problem.

Future research will aim to explore prospective directions by applying our proposed data augmentation method with other deep learning architectures, such as AlexNet, ResNet101, and DenseNet201, for efficient detection of skin melanoma disease. Secondly, the proposed image augmentation method can be applied to other imaging domain such as in plant disease detection, fault detection, etc. to improve the existing research bottleneck in small data analytics. Therefore, we will test our method on other datasets. Finally, we hope to integrate our model into a smartphone application for developing the AI based decision support system to aid user's self-clinical examination thus contributing to early detection of skin cancer diseases.

References

[1] Codella NC, Nguyen QB, Pankanti S, Gutman DA, Helba B et al. Deep learning ensembles for melanoma recognition in dermoscopy images. IBM Journal of Research and Development. 2017; 61 (4/5):5–1.

[2] Ahmed H, Al-azawi R, Abdulhameed A. Evaluation methodology between globalization and localization features approaches for skin cancer lesions classification. In: Journal of Physics: Conference Series. vol. 1003; 2018. p. 012029.

[3] Barata C, Ruela M, Francisco M, Mendonca T, Marques JS. Two systems for the detection of melanomas in dermoscopy images using texture and color features. IEEE Systems Journal. 2013; 8 (3):965–979.

[4] Brinker TJ, Hekler A, Enk AH, Berking C, Haferkamp S et al. Deep neural networks are superior to dermatologists in melanoma image classification. European Journal of Cancer. 2019;119:11–17.

[5] Ratul MAR, Mozaffari MH, Lee W, Parimbelli E. Skin Lesions Classification Using Deep Learning Based on Dilated Convolution. bioRxiv. 2020:860700.

[6] Oh BH, Kim KH, Chung KY. Skin Imaging Using Ultrasound Imaging, Optical Coherence Tomography, Confocal Microscopy, and Two-Photon Microscopy in Cutaneous Oncology. Frontiers in Medicine. 2019;6:274.

[7] Barata C, Celebi ME, Marques JS. A survey of feature extraction in dermoscopy image analysis of skin cancer.IEEE journal of biomedical and health informatics. 2018; 23 (3):1096–1109.

[8] Li H, He X, Zhou F, Yu Z, Ni D et al. Dense deconvolutional network for skin lesion segmentation. IEEE Journal of Biomedical and Health Informatics. 2018; 23 (2):527–537.

[9] Maiti A, Chatterjee B. Improving detection of Melanoma and Naevus with deep neural networks. Multimedia Tools and Applications. 2019:1–20.

[10] Zhang J, Xie Y, Xia Y, Shen C. Attention residual learning for skin lesion classification.IEEE transactions on medical imaging. 2019; 38 (9):2092–2103.

[11] Hosny KM, Kassem MA, Foaud MM. Classification of skin lesions using transfer learning and augmentation with Alex-net. PloS one. 2019; 14 (5):e0217293.

[12] Pacheco AG, Krohling RA. The impact of patient clinical information on automated skin cancer detection. Computers in biology and medicine. 2020;116:103545

[13] Curiel-Lewandrowski C, Berry E, Leachman S. Artificial intelligence approach in melanoma. Melanoma New York (NY): Springer. 2019:1–31.

[14] Lateh MA, Muda AK, Yusof ZIM, Muda NA, Azmi MS. Handling a Small Dataset Problem in Prediction Model by employ Artificial Data Generation Approach: A Review. Journal of Physics: Conference Series. 2017 Sep;892:012016.

[15] Kassani SH, Kassani PH. A comparative study of deep learning architectures on melanoma detection. Tissue and Cell. 2019; 58:76–83.

[16] Nasr-Esfahani E,Rafiei S,Jafari MH,Karimi N, Wrobel JS et al. Dense pooling layers in fully convolutional network for skin lesion segmentation.Computerized Medical Imaging and Graphics. 2019;78:101658.

[17] Hekler A,Utikal JS,Enk AH,Berking C, Klode J et al. Pathologist-level classification of histopathological melanoma images with deep neural networks. European Journal of Cancer. 2019;115:79–83.

[18] Zhang C, Wang X, Zhang F, Dong L, Wu J, Cheng Q, et al. Phenylalanine ammonia-lyase contributes to the soybean response towards Phytophthora sojae infection. Scientific reports. 2017; 7 (1):1–13.

[19] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research. 2002;16:321–357.

[20] Khan MA, Akram T, Sharif M, Javed K, Rashid M et al. An integrated framework of skin lesion detection and recognition through saliency method and optimal deep neural network features selection. Neural Computing and Applications. 2019:1–20.

[21] Abayomi-Alli OO, Damaševičius R, Wieczorek M, Wozniak M. Data Augmentation Using Principal Component Resampling for Image Recognition by Deep Learning. In: International Conference on Artificial Intelligence and Soft Computing. Springer; 2020. pp. 39–48.

[22] Urbonas A, Raudonis V, Maskeliunas R, Damasevicius R. Automated identification of wood veneer surface defects using faster region-based convolutional neural network with data augmentation and transfer learning. Applied Sciences. 2019; 9 (22):4898.

[23] Leguen-deVarona I, Madera J, Martinez-Lopez Y, Hernandez-Nieto JC. SMOTE-Cov: A New Oversampling Method Based on the Covariance Matrix. In: Data Analysis and Optimization for Engineering and Computing Problems.Springer; 2020. pp. 207–215.

[24] Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size; 2016.

[25] Munir K, Elahi H, Ayub A, Frezza F, Rizzi A.Cancer diagnosis using deep learning: a bibliographic review. Cancers. 2019; 11 (9):1235.

[26] Roda A, Michelini E, Zangheri M, Di Fusco M, Calabria D et al. Smartphone-based biosensors: A critical review and perspectives. TrAC Trends in Analytical Chemistry. 2016;79:317–325.

[27] Rahimzadeganasl A, Alganci U, Goksel C. An Approach for the Pan Sharpening of Very High Resolution Satellite Images Using a CIELab Color Based Component Substitution Algorithm. Applied Sciences. 2019; 9 (23):5234.

[28] Qiu J, Harold Li H, Zhang T, Ma F, Yang D. Automatic x-ray image contrast enhancement based on parameter auto-optimization. Journal of applied clinical medical physics. 2017; 18 (6):218–223.

[29] Kim SJ, Min BS, Lim DK, Lee JH. Determining parameters in contrast limited adaptive histogram equalization. In: Proceedings of the 7th International Conference on Information Security and Assurance. vol. 21; 2013. pp. 204–207.

[30] Frangi AF, Niessen WJ, Vincken KL, Viergever MA. Multiscale vessel enhancement filtering. In: International conference on medical image computing and computer-assisted intervention. Springer; 1998. pp. 130–137.

[31] Mumford DB, Shah J. Optimal approximations by piecewise smooth functions and associated variational problems. Communications on pure and applied mathematics. 1989.

[32] Sahlol AT,Abd Elaziz M, Tariq Jamal A,Damasevicius R,Farouk Hassan O. A novel method for detection of tuberculosis in chest radiographs using artificial ecosystem-based optimisation of deep neural network features. Symmetry. 2020;12 (7):1146.25

[33] Mendonca T,Ferreira PM,Marques JS,Marcal AR,Rozeira J. 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). IEEE; 2013. pp. 5437–5440.

[34] Khan MA, Javed MY, Sharif M, Saba T, Rehman A. Multi-model deep neural network based features extraction and optimal selection approach for skin lesion classification. In: 2019 international conference on computer and information sciences (ICCIS). IEEE; 2019. pp. 1–7.

[35] Hu K, Niu X, Liu S, Zhang Y, Cao C, Xiao F, et al. Classification of melanoma based on feature similarity measurement for codebook learning in the bag-of-features model. Biomedical Signal Processing and Control. 2019; 51:200–209.

[36] Adjed F, Gardezi SJS,Ababsa F,Faye I,Dass SC. Fusion of structural and textural features for melanoma recognition. IET Computer Vision. 2017; 12 (2):185–195.

[37] Eltayef K, Li Y, Liu X. Detection of melanoma skin cancer in dermoscopy images. In: Journal of Physics: Conference Series. vol. 787. IOP Publishing; 2017. p. 012–034.

[38] Bi L, Kim J, Ahn E, Feng D, Fulham M. Automatic melanoma detection via multi-scale lesion-biased representation and joint reverse classification. In: 2016 IEEE 13th international symposium on biomedical imaging (ISBI). IEEE; 402016. p. 1055–1058.