Skin Cancer Detection Using Ensembled Support Vector Machines

Abhishek Das[1] and Mihir Narayan Mohanty[2,*]
[1] Department of CSE, Centurion University of Technology and Management, Paralakhemundi, Odisha, India
[2]ITER, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India
[1]abhishekdas225@gmail.com and [*2]mihir.n.mohanty@gmail.com

Abstract: Manual skin cancer detection from dermatoscopic images is difficult due to the similar appearance among different classes of skin cancer. Applying machine learning techniques can provide faster and more accurate predictions that will assist physicians in easy diagnosis. The class imbalance problem in the dataset is verifid and overcome with SMOTE-TOMEK and target-specific augmentation methods. Augmentation is done by rotating, flipping, shifting, zooming, and varying the brightness of the images. The ultimate effort is given to design an appropriate classifier that receives the raw images as input and accurately predicts. A homogeneous ensemble learning-based model designed with three support vector machines (EnsembleSVM) is designed to detect skin cancer. The ensemble model consists of two stages. In the first stage, two SVM models are connected in parallel. One model is trained with the balanced dataset generated using SMOTE-TOMEK, whereas the second SVM model is trained with the augmented data. The final classification is done by another SVM model stacked to the ensemble of two SVMs. The proposed algorithm is verified using the HAM10000 dataset. The proposed method analyses the effect of class imbalance problem and provides a solution to it. Appropriate choice of clasifyig model in a better way, i.e. ensembleing the models, provided 99.9% training accuracy and 98% test accuracy. It shows the effectiveness of the proposed method in detecting skin cancer from dermatoscopic images.

Keywords: Skin Cancer Detection, Biomedical Image Classification, Ensemble Learning, SMOTE-TOMEK, Data Augmentation, Support Vector Machine

Skin Cancer Detection Using Ensembled Support Vector Machines

Abstract: Manual skin cancer detection from dermatoscopic images is difficult due to the similar appearance among different classes of skin cancer. Applying machine learning techniques can provide faster and more accurate predictions that will assist physicians in easy diagnosis. The class imbalance problem in the dataset is verifid and overcome with SMOTE-TOMEK and target-specific augmentation methods. Augmentation is done by rotating, flipping, shifting, zooming, and varying the brightness of the images. The ultimate effort is given to design an appropriate classifier that receives the raw images as input and accurately predicts. A homogeneous ensemble learning-based model designed with three support vector machines (EnsembleSVM) is designed to detect skin cancer. The ensemble model consists of two stages. In the first stage, two SVM models are connected in parallel. One model is trained with the balanced dataset generated using SMOTE-TOMEK, whereas the second SVM model is trained with the augmented data. The final classification is done by another SVM model stacked to the ensemble of two SVMs. The proposed algorithm is verified using the HAM10000 dataset. The proposed method analyses the effect of class imbalance problem and provides a solution to it. Appropriate choice of clasifyig model in a better way, i.e. ensembleing the models, provided 99.9% training accuracy and 98% test accuracy. It shows the effectiveness of the proposed method in detecting skin cancer from dermatoscopic images.

Keywords: Skin Cancer Detection, Biomedical Image Classification, Ensemble Learning, SMOTE-TOMEK, Data Augmentation, Support Vector Machine

1. Introduction

According to the report in [1], the estimated number of new cases of melanoma skin cancer in 2020 was 324635, whereas 57043 people died in that year out of skin cancer. Faster and more accurate skin cancer detection may lead to a decrease in this count. Clinical manual screening of dermatoscopic or biopsy is time-consuming and not that easy due to the irregular shapes and heterogeneous appearance of skin lesions. The difficulty for early remedies is the high false-negative incidence of malignant melanoma due to skin lesions' varying forms and variable appearance. Current techniques are progressing to overcome the time-consuming traditional methods with the help of machine learning.

Skin cancer detection from dermatoscopic images has been studied in various forms. Preprocessing steps have importance in most of the works. Works developed in this field of research are discussed in this section for analysis purposes. Various steps followed in different state-of-art methods have been discussed in [2]. From that review article, one can observe that different measures have been followed for skin cancer detection from dermatoscopic images. Hair removal, noise elimination, segmentation of the lesion portion from the whole image, image normalization, feature extraction, feature selection, feature optimization, training, and classification are the basic steps in skin cancer detection. ABCD (Asymmetry, Border, Color, and Diameter) features [3-5] are mostly preferred feature sets for analyzing skin dermatoscopic images. Gray Level Co-Occurrence Matrix (GLCM) features [6-8] represent the dermatoscopic images used to train the machine learning models. Statistical feature extraction before training helps in performance improvement, whereas it adds extra computational cost and time consumption. In most of the works, it is habitual that traditional machine learning models such as Support vector machine (SVM), K-nearest neighbor (KNN), and random forest (RF) [9] have not been fed directly by the data but by various statistical features. At the same time, deep learning models have been fed by direct data due to their auto-feature extraction tendency. Proper model selection and the combination can overcome all these perceptions and makes the detection accurate [10,11]. The models developed using machine learning techniques are

primarily used in the single form, whereas hybrid forms must be explored more. Learning models that provide single predictions suffer from three fundamental problems, and are as follows.

- Lack of statistical analysis: a hypothesis created by the trained single model may not classify the test data correctly.
- Computational problem: when the learning algorithm does not ensure that it will find the optimal hypothesis inside the hypothesis space.
- Representation problem: A representation problem arises when a particular solution is absent from any test data.

Ensemble learning is one of the alternate solutions to these problems. Ensemble learning-based models are proven to be a better choice in biomedical image classification. It has been applied for skin cancer detection [12] from dermatoscopic images, as well as brain tumor detection [13] from magnetic resonance images, breast cancer detection [14] from breast histopathology images, and many more. In this work, SVM, the popular supervised machine learning model, is used for classification in an ensemble fashion and named EnsembleSVM. Although the RBF kernel is more widespread [15] in comparison to other kernels of SVM, the polynomial kernel worked well with the dataset used in this work. Instead of extracting statistical features from the input images, the images are directly fed to the proposed model, and the result obtained is satisfactory. The highlights of this work are summarized as follows:

- The class imbalance problem is overcome with SMOTE-TOMEK and the target-specific data augmentation technique.
- Instead of extracting any statistical or structural features, the images are directly fed to the ensemble of three SVMs termed EnsembleSVM to simplify the process.
- The EnsembleSVM model's performance is improved compared to the recent deep learning models.

Research in the field of skin cancer detection includes different forms of learning algorithms. Classical approaches are found in recent works. The hidden Markov Model (HMM) has been used for skin cancer detection from dermatoscopic images [16]. In that work, the authors optimized the parameters of the HMM model using expectation maximization (EM), and asymmetric textures were studied using asymmetric analysis. Application of Histogram of Gradients (HG) and Histogram of Lines (HL) has been verified with different textural and color features for dermatoscopic image classification [17]. These methods performed better than the standard Histogram of Oriented Gradient (HOG) and Histogram of Oriented Lines (HOL) methods. The statistical feature extraction, including color vector angles and Zernike Moments (ZM), makes that work different from others.

Numerous machine learning-based models and different feature extraction methods for detecting skin cancers from dermatoscopic images have been designed. A multi-tree programming model has been proposed for melanoma detection [18]. The authors have followed feature construction, feature selection, training, and classification steps in that work. Color information was extracted using a Local binary pattern (LBP).Global features like color variation within lesions, among lesions, and border shapes were considered for the training of six different classifiers as Naïve Bayes (NB), SVM with radial basis function (RBF) kernel, k-NN with k=5, decision trees (DT), Multilayer Perceptron (MLP), and random forest (RF). Another work has been developed by considering hair removal from dermatoscopic images before applying any classification strategy [19]. In that work, the authors have used the maximum gradient intensity algorithm for hair removal. Segmentation was adopted for lesion extraction from the image using the Otsu thresholding algorithm. Features were extracted using the ABCD, Gray Level Co-Occurrence Matrix (GLCM), and LBP. Machine learning models such as KNN, SVM, and Decision trees were employed for skin cancer detection from dermatoscopic images [20]. Hair removal, noise reduction, segmentation by cropping, and binarization of the skin cancer lesion, followed by textural, shape, and color feature extraction and feature optimization using the PSO variant, were adopted as preprocessing steps.

Deep learning-based models have occupied the highest count in recent decades with auto feature extraction and detection strategies. Sequential dermatoscopic images have been used for melanoma

detection [21]. In that work, the sequential images were considered to observe the change in lesion growth, and a classification model was designed to detect cancer types. Estimated Euclidean transformation was used to align the images in a sequence, then the region of lesions was extracted, followed by the training and classification by the Spatio-Temporal Network (STN). The application of reinforcement learning for segmentation has been addressed for skin image classification from dermatoscopic images [22]. The model used for segmentation was designed using an encoder on the input side and a decoder on the output side. A neural network search algorithm using a hill-climbing algorithm has been proposed in an article [23] for skin lesion classification from dermatoscopic images. The search method was optimally working on designing an ensemble model using convolutional layers for best performance. A multi-task deep learning model that can perform detection, classification, and segmentation at a time has been proposed in [24] for skin lesion analysis from dermatoscopic images. The loss function was calculated using focal loss and the Jaccard distance. Image resize, zero normalization, and data augmentation was used as data preprocessing steps. The augmented images were used to train the deep learning framework. The feature pyramid network (FPN) was used to enhance the features, followed by a region proposal network (RPN) for generating the region of interest. Three convolutional neural networks were used to process these data for detection, classification, and segmentation in parallel. An encoding-decoding model has been designed using convolutional layers for melanoma and non-melanoma classification from dermatoscopic images [25]. The preprocessing steps, like noise reduction using a Gaussian filter, resizing the images to 224x224, normalization, and augmentation, were adopted before passing through the deep learning model. The classification was done in two levels, one in pixel-level classification using the Softmax classifier and the second one for lesion classification in the lesion classifier. Melanoma detection uses a lightweight CNN model with feature discrimination from dermatoscopic images [26]. In that work, the authors have applied semantic segmentation to extract the skin lesion from the input images. Features from the positive and negative images were fed to the classification model and the feature discrimination networks at a time to improve the performance. Polarization speckle images have been utilized to classify skin cancer types from dermatoscopic images with various machine learning and deep learning models [27]. The works have involved the data augmentation step in increasing the data size. Gamma correction was applied to the polarization speckle images as another preprocessing step. The preprocessed augmented data were fed to the ResNet model. A deep Q net model optimized with fractional student psychology optimization (FSPO) technique has been proposed for skin cancer detection [28]. That method included noise removal using a Type II fuzzy system optimized by Cuckoo search optimization. Segmentation was then applied to the noise-free images using a speech-enhanced generative adversarial network (SeGAN). The segmented images were passed through FSPO-optimized deep Q Net for training and classification. An extreme learning machine (ELM) has been used for skin cancer detection from dermatoscopic images [29]. The authors in that work have passed the input images through different preprocessing steps like contrast enhancement and feature extraction using CNN. Feature selection has been made using a whale optimization algorithm combined with entropy mutual information (EMI). Features were combined using the canonical correlation method. Concatenation of a convolutional neural network with SVM has been proposed in [30] for skin cancer detection. The hyperparameters of CNN and SVM models were optimized using the Bayesian optimization algorithm (BOA). An ensemble of different forms of CNN models has provided 89% accuracy with Resnet and Inception V3 as base models [31], 91% accuracy with Resnet and Inception [32], and a maximum of 96.57% F1-score with the ensemble of Resnet, DenseNet, SE-ResNext, and NASNet [33].

The rigorous training makes the performance accurate. It is observed that all the classical models and statistical machine learning models have been designed with feature extraction steps. Authors who have used deep learning models as classification models have avoided using any preprocessing steps; however, a few works have applied preprocessing steps. Ensemble learning provides better accuracy in comparison to single models. Different preprocessing measures are considered in most of the works, even with the ensemble models. A balanced dataset enhances the performance of a classifier [34-36]. However, the class imbalance in dermatoscopic datasets has yet to be considered in state-of-art methods. This work addresses the class imbalance problem, and its effect on classification performance is verified with the ensemble

learning-based decision-making technique. The use of feature extraction steps and other preprocessing steps is not considered even if the classifying models belong to the statistical machine learning category. The results obtained by the proposed model are competitive with the deep learning models.

The rest of the part is organized as follows. The proposed method is described elaborately in Section 2. Results obtained by the proposed model are provided and discussed in Section 3. The work is discussed in Section 4. Section 5 concludes the work at a glance and offers future perspectives on this work.

2. Methodology

Class imbalance problem in the dataset is handled with SMOTE-TOMEK and data augmentation. The ensemble model is formed with two base models and a single meta classifier. Deep learning models are well utilized when the dataset is significant for training and classification. A single SVM can not be suitable for a large dataset. In this context, the design of a model with ensemble learning is proposed in this work with three SVMs. The ensemble learning-based models perform better than single models [37]. Ensemble learning-based models are broadly classified into two groups; one is heterogeneous, and the other one is homogeneous type. Heterogeneous models have different base classifiers, whereas homogeneous models are developed using base classifiers of the same learning algorithm. The homogeneous type of ensemble learning method is used with all the base models of the same learning algorithm, i.e., SVM with a polynomial kernel of degree 3. As the base models are of the same configuration, training is performed with two different forms of data to make the overall training robust. One model is trained with the direct dataset, whereas the second one is trained with augmented data. The SVM does the final classification with the polynomial kernel trained by the base models' predictions. The workflow diagram of the proposed method is shown in Fig. 1.

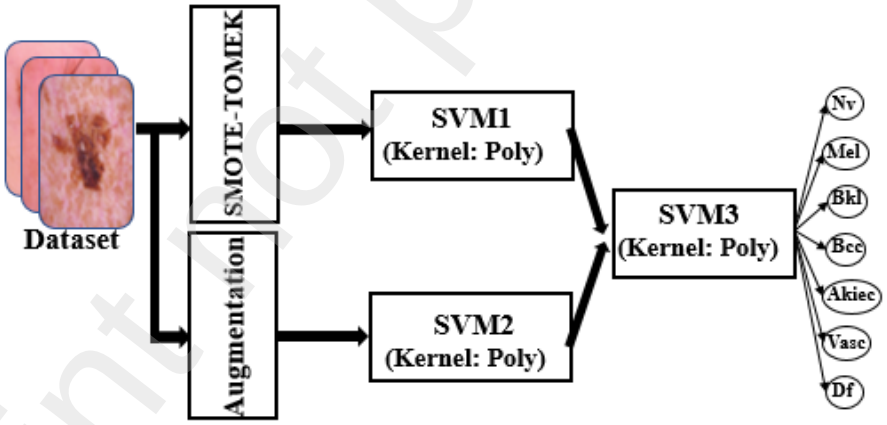


Fig. 1. Block Diagram of the proposed work

2.1. Dataset Description

The HAM10000 dataset is used in this work, which is publicly available on Kaggle [38]. The dataset contains a total of 10015 dermatoscopic images of seven categories, i.e., melanocytic nevi (Nv), melanoma (Mel), benign keratosis-like lesions (Bkl), basal cell carcinoma (Bcc), actinic keratoses (Akiec), vascular lesions (Vasc), and dermatofibroma (Df) with 6075, 1113, 1099, 514, 327, 142, and 115 images respectively. The 28x28 RGB images are considered; hence each image has three channels. The whole dataset is divided into train, and test sets with 80:20 ratios, i.e., train and test sets contain 8012 and 2003 images, respectively. Sample images of each category of skin cancer are shown in Fig. 2.

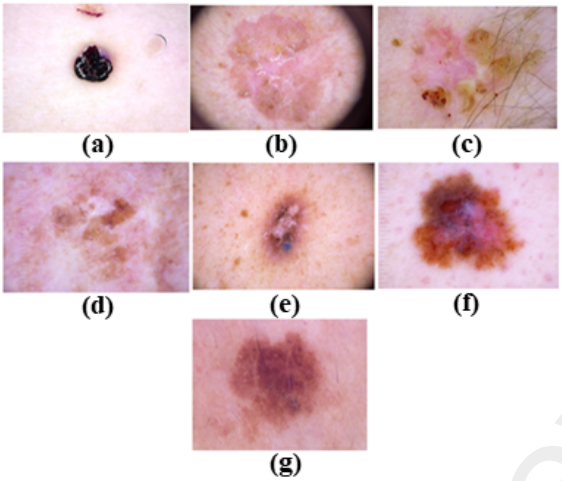


Fig. 2. Sample images from every seven classes (a) Vascular lesions, (b) Actinic Keratoses, (c)Basal cell carcinoma, (d) Benign keratosis-like lesions, (e) Dermatofibroma, (f) Melanoma, (g) Melanocytic nevi

2.2. SMOTE-TOMEK Links

SMOTE-Tomek Links was introduced first by Batista et al. [39]; that combines the SMOTE ability to generate synthetic data for the minority class and Tomek Link's ability to remove the data that are identified as Tomek links from the majority class (that is, samples of data from the majority class that is closest with the minority class data). Once the synthetic data points are formed using SMOTE, random data from the majority class are chosen. These points are checked for overlapping with any data points from the minority class and form the TOMEK Links to remove the overlapped points.

The SMOTE step begins by determining the number of nearest neighbors (k), then calculating the shortest distance between the random data selected from the minority class and the data of the k-nearest neighbors using the Euclidean distance. Furthermore, based on the closest distance, the synthetic sample data are generated for the minority class. The process is stopped when the data for each class have been balanced. The first step in Tomek Links is choosing a pair of samples with the minimum Euclidean distance from the k-nearest neighbors, where each sample comes from a different class. The sample pair is the Tomek Link, if no sample satisfies the following conditions related to the Euclidean distance greater or less than that of the other class.

2.3. Data Augmentation

There is a significant variation among the classes regarding the number of images. This variation may lead to improper training of machine learning models. Hence, data augmentation is applied to each type with fewer images compared to the class with the highest count in the train set, keeping the test set the same after partition.

The class Nv in the train set has 5378 images which are the highest among all the types, whereas Mel, Bkl, Bcc, Akiec, Vasc, and Df have 891,871,418,266,110 and 78 images, respectively. This count variation was overcome by applying data augmentation in these classes to have 5378 images in each class.

Data augmentation is done by shifting the images in horizontal and vertical directions, flipping the images horizontally and vertically, and rotating the images at different angles, i.e., 30^0, 45^0, 60^0, and 90^0. Other data augmentation steps, such as brightness variation, are utilized within the range of 0.2 and 0.9 to avoid data loss, and zooming is applied to images within the range [0.5,1.0]. The results obtained from these steps are shown in the result section. This process was applied to the low-count classes until it reached

5378. The augmentation steps followed to design a target-based augmentation (TAug) model are provided in Algorithm 1.

After applying the data augmentation, the augmented train set containing 37646 images with 5378 images in each class is obtained. A comparative analysis of the effect of augmentation on training and training with direct data is provided in the result section.

Algorithm 1: Designing Target-based Augmentation

1. Input: Dataset D $= \{\mathbf{x_i}, y_i\}_{i=1}^{n}$
2. Output: D_{aug}

3. **Step 1:** Read the data 'D'
4. **Step 2: Class size evaluation**
5. $y = categorical(y)$
6. $c = counter(y)$
7. $t = max(c)$
8. **Step 3: Import augmentation model (TAug)**
9. For $i = 1\ to\ 6$
    a. $D_{aug_i} = \boldsymbol{TAug}\ (\boldsymbol{D_i})$
    b. **If** $\boldsymbol{size}(D_{aug_i}) = \boldsymbol{t}$
    c. **then Stop**
10. End For
11. Aug_D = Merge(D_{aug},D_0)

2.4. Ensemble Model Design

2.4.1. Support Vector Machine (SVM)

SVM is one of the powerful tools in machine learning whose strategy is to separate the classes using hyperplanes. SVM models support kernels that add more features to the data to make them easily detachable. Gaussian kernel (also known as RBF kernel) maps the large-size input data to an infinite-dimension feature space that decays faster than that of small-size data. It results in a quadratic increase in memory requirement to store the features, and hence time consumption also increases quadratically. Polynomial kernel with deficient degree performs like linear kernels and cannot separate the nonlinearly distributed classes. The degree of the polynomial kernel affects the classifier's performance, and higher-degree increases the flexibility of the classifier to separate the classes [40-42] properly. This motivated us to choose SVM with a polynomial kernel of degree 3 to classify high-dimensional data with seven nonlinearly separated classes.

This work involves the dataset whose class distribution is found to be nonlinear, as shown in Fig. 3. Both the polynomial and RBF nonlinear kernels are verified, and it is found that the polynomial kernel fits best with the data. This has been verified in earlier works that for nonlinear distribution of data, polynomial kernels with a higher degree have resulted in better classification [43,44]. The data regularization is not required when the kernel with a higher order degree is applied, and it is similar to training over-parameterized ReLU networks without regularization. The results obtained using SVM with RBF kernel are provided in the result section compared to the polynomial kernel.

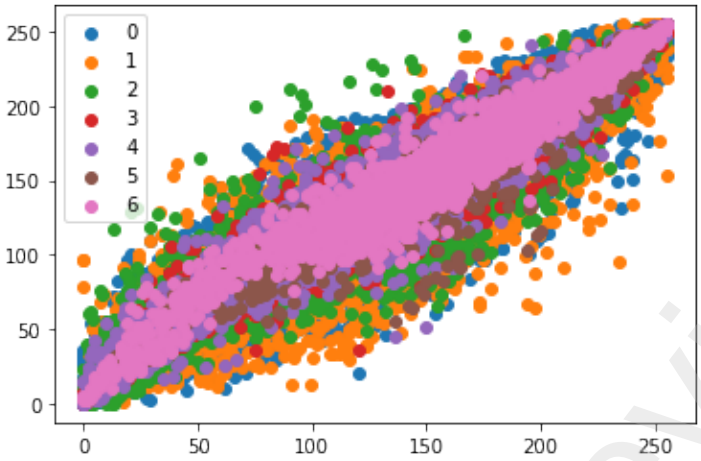


Fig. 3. The scatter plot of the skin cancer dataset

In Fig. 3, the seven classes are shown with seven different colors. The classes are nonlinearly distributed as there is overlap among the classes.

2.4.2. Stacking

In the stacked generalization, the R^n space occupied by the original data set D is termed base space, and the models used as generalizers at this stage are termed the base classifier. The base classifiers take the input data points to a new space R^k. The output of each base classifier contains the corresponding predictions on the input data (x_i,y_pred_i). It forms the training data for the next level of generalization, i.e., stacking another model for the final prediction.

A homogeneous ensemble learning-based model with two base classifiers and one meta-classifier is proposed. The base models are directly trained with the two forms of the data, whereas the final classifier (termed meta classifier) is trained with the outcomes (y_pred_1,y_pred_2) of the base classifiers SVM1 and SVM2. A parallel connection is adopted for parallel training of the base models, and their decisions are then passed to the meta-classifier SVM3 for the final decision. Hence, the output of the stacked meta classifier SVM3 is a function of y_pred_1 and y_pred_1.

$$y_{output}(\mathbf{x_i}) = SVM3(y\_pred_{i1},y\_pred_{i2}) \quad (1)$$

Where x_i represents the input to the base classifiers.

The proposed method is developed using Algorithm 2.

Algorithm 2: Training of the proposed model

1. Input: Dataset $D = \{\mathbf{x_i},y_i\}_{i=1}^{n}$
2. Output: y_{output}

3. **Step 1:** SMOTE-TOMEK Links (STL)
4. $D_{STL} = STL(D)$
5. **Step 2: Augmentation**
6. Import the augmentation model (T_Aug)
7. $D_{aug} = T\_Aug(D)$
8. **Step 3: Train the Base Classifiers**
9. Import base classifiers (BC)

10. Train BC_1 with D_{STL}
11. Train BC_2 with D_{aug}
12. **Step 4:** Generate the input data for Meta Classifier
13. $For\ i = 1\ to\ n$
    $and\ j = 1\ to\ m$ do
    D_{meta}=$\{\mathbf{x'_{ij}}, y_i\}$,

    where x'_{ij}= Concatenation ($BC_1(\mathbf{x_{STLi}})$, $BL_2(\mathbf{x_{aug_j}})$)
14. end For
15. **Step 5: Train Meta Classifier (MC)**
16. Train MC with D_{meta}
17. Return y_{output}

Where n represents the number of images in the original dataset, and m represents the number of augmented images.

3. Results

3.1. Image Augmentation

The class imbalance problem in the dataset is handled using the proposed target-oriented data augmentation method. The outputs of different augmentation sub-strategies for a sample input image are shown in Fig. 4 and Fig. 5.

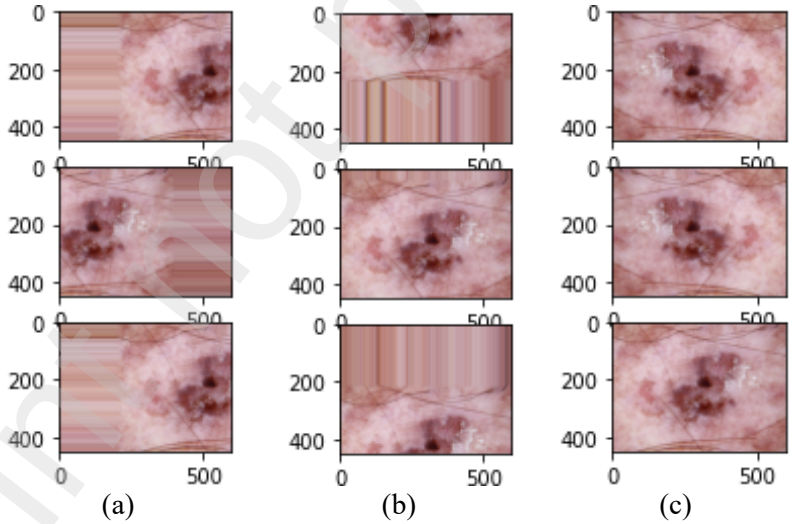


Fig. 4. Outputs after applying (a) horizontal shift, (b) vertical shift, (c) horizontal and vertical flip

A few images are generated by shifting and flipping the images in different directions to create new data to balance the classes. Fig. 4 shows the sample of one input image going through various stages of shifting and flipping.

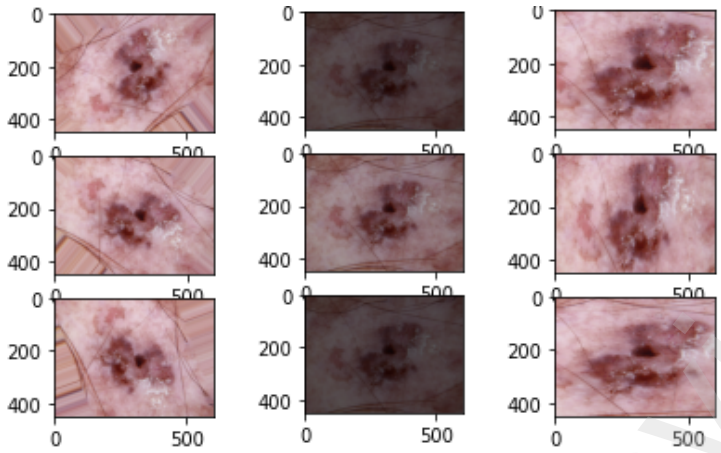


Fig. 5. Outputs after applying (a) rotation, (b) brightness variation, (c) zooming

Fig. 5 shows the generation of new data from a sample input image by rotating it in different directions. New images are generated by varying the brightness and zooming out and zooming the images.

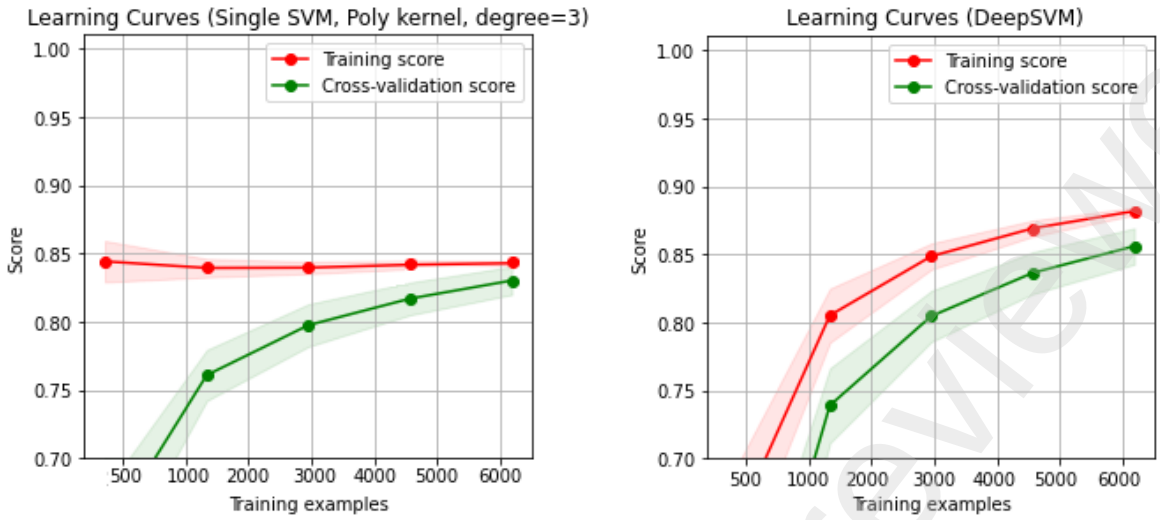
3.2. Performance Evaluation

At the initial stage, the dataset is applied to SVM with RBF kernel as it is the most preferred kernel. The result obtained from this model is shown in Table 1. The SVM-RBF model provided 78.68% training accuracy and 73.79% testing accuracy, very low values of F1-score, precision, and recall. The performance obtained in this stage is not higly accurate and needs further improvement. The numerical values 0 to 6 represent seven different classes of skin cancer present in the dataset.

Table 1. Performance of the SVM-RBF model on imbalanced dataset

	Precision	**Recall**	**F1-Score**	**Support**
0	0.78	0.96	0.86	1374
1	0.51	0.13	0.20	205
2	0.55	0.26	0.35	227
3	0.51	0.49	0.50	94
4	0.38	0.22	0.28	55
5	0.61	0.39	0.48	28
6	0.50	0.05	0.09	20
Accuracy			0.74	2003
Macro Avg	0.55	0.36	0.39	2003
Weighted Avg	0.70	0.74	0.69	2003

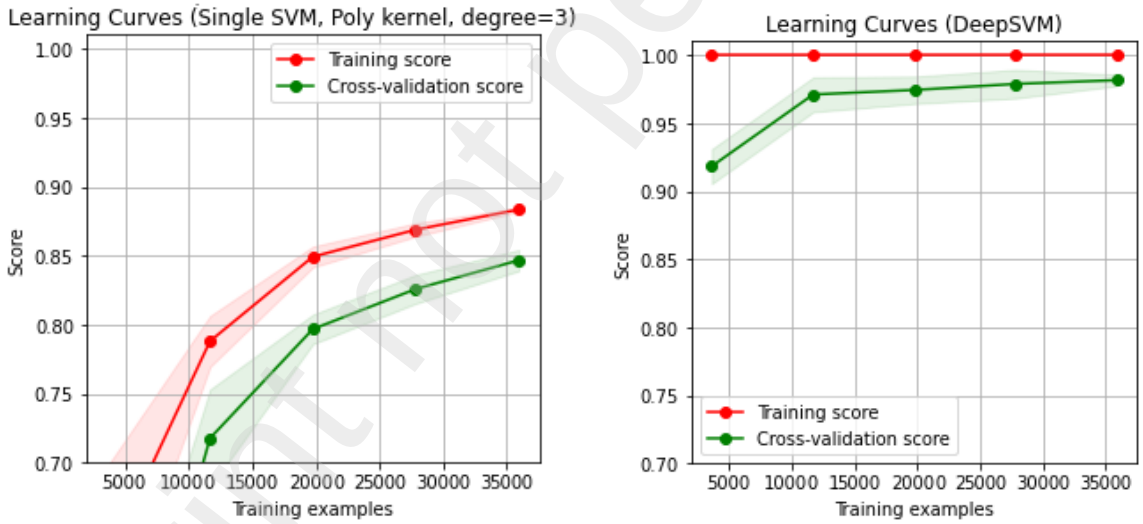
Further, polynomial kernel is chosen, and verified the effect of original data on the performance of the single model of SVM, as well as with the proposed EnsembleSVM model. The ensemble model took 13 minutes to fully train on the original and balanced datasets. The trained model predicts a single test data in fractions of second. A fully trained model takes 1.27 Megabytes for storage. The results obtained with the original images were insufficient for accurately detecting skin cancer. The results are shown in Fig. 6.



(a) (b)

Fig.6. Performance obtained by using original data with (a) a single SVM model and (b) an EnsembleSVM model

Such low performance was observed due to class imbalance which was verified by balancing the classes with the application of data augmentation. The single and proposed EnsembleSVM models proved the augmented and balanced dataset. The results regarding training and validation scores with 5-fold cross-validation are shown in Fig. 7.



(a) (b)

Fig. 7. Performance obtained by using balanced data with (a) a single SVM model and (b) an EnsembleSVM model

The explanation of the results from Fig.6 and 7 are given in Table 2 and discussed in the Discussion subsection. The training, validation, and testing score for each combination of data and model, i.e., whether the direct data is applied to a single SVM or EnsembleSVM and the augmented data is used to a single SVM or EnsembleSVM, are provided in Table 3.

Table 2. The performance metrics of each model

		Training Score (in %)	Validation Score (in %)	Testing Score (in %)
Direct Data	Single SVM	84	82.6	80.38
	EnsembleSVM	88.72	85.76	83.13
Augmented Data	Single SVM	88.53	84.93	82.28
	EnsembleSVM	**99.9**	**98**	**98**

From Table 2, it is observed that the stacked ensemble models are performing better in comparison to the single model. Data augmentation is affecting performance positively. The proposed EnsembleSVM model provides the highest scores in training and validation with the augmented form of the dataset.

The model's performance is verified by applying the earlier test set for all the models trained with single and augmented data. The results obtained from the test set in terms of confusion matrices are shown in Fig. 8.

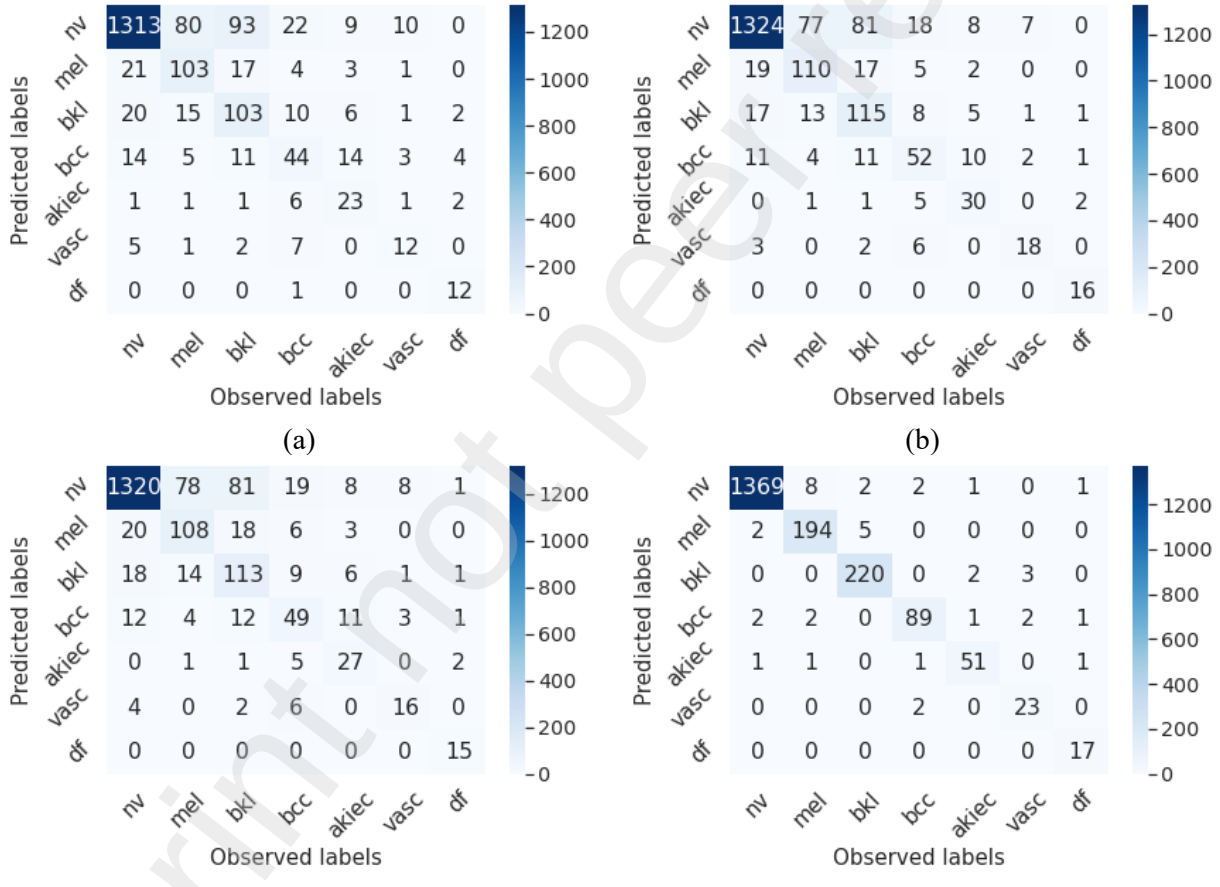


Fig. 8. Confusion matrix for (a) single SVM trained with the direct dataset, (b) EnsembleSVM trained with the dataset, (c) single SVM trained with the augmented dataset, and (d) EnsembleSVM trained with an augmented dataset

Confusion matrices in Fig. 9 are formed using the 2003 test data numbers formed at the original dataset's beginning. This set is used as the test set for all the models trained with different forms of data. The confusion matrix calculates the proposed model's accuracy, F1 score, precision, and recall. It is found that

the proposed EnsembleSVM model trained with augmented data provides the highest testing accuracy. The accuracy and other parameters were evaluated using Eqn. (2-5) Furthermore, they are given in Table 3.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \times 100 \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \times 100 \quad (4)$$

$$F1\ Score = 2 \times \frac{Recall \times Precision}{Recall + Precision} \quad (5)$$

Table 3. Evaluating parameters for EnsembleSVM on the test data

	Support	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Nv	1374	99.05%	0.99	1.0	0.99
Mel	205	99.1%	0.97	0.95	0.96
Bkl	227	99.4%	0.98	0.97	0.97
Bcc	94	99.35%	0.92	0.95	0.93
Akiec	55	99.6%	0.93	0.93	0.93
Vasc	28	99.65%	0.92	0.82	0.87
Df	20	99.85%	1.0	0.85	0.92

The comparison analysis of the proposed work with the *state-of-art* models is provided in Table 4.

Table 4. Comparison of the proposed model with *state-of-the-art models*

Reference	Preprocessing	Feature extraction	Classification	Testing Accuracy (in %)
[17]	-	Color features and Location features	SVM-RBF	93
[18]	-	-	DenseNet-169	92.25
[19]	Contrast enhancement	CNN features	ELM	93.4
[20]	Segmentation	-	EncoderDecoder FCN	95.33
[21]	-	-	Optimized Ensemble Model Using Convolutional Layers	77
[22]	-	-	BOA optimized (CNN+SVM)	90
[23]	Segmentation using fuzzy kernel C-means	-	Red fox optimization-based MLP	90.5
[24]	-	Textural and color features	HG and HL	92.96
[25]	-	-	Cloud-based CNN	96.27
[26]	-	ABCD	Total Dermatoscopic Score	84
[27]	Hair removal, smoothing, noise removal, edge detection	ABCD and GLCM	SVM	96.25

[28]	Augmentation and Gamma correction	-	ResNet	82
[29]	Noise reduction, resizing, normalization, and augmentation	-	CNN-based EncoderDecoder	95
[30]	Semantic segmentation	-	Ensemble of 2 CNNs	96.2
[31]	-	-	Ensemble of Resnet and Inception V3	89
[32]	-	-	Ensemble of Resnet and Inception	91
[33]	-	-	Ensemble of Resnet, DenseNet, SE-ResNext, and NASNet	96.57
This work	SMOTE-TOMEK Links and target-based Augmentation for class balancing	-	EnsembleSVM (Ensemble of 3 SVMs)	**98**

4. Discussion

This work is started with the aim of detecting skin cancer from dermatoscopic images keeping in the mind to reduce the preprocessing steps, developing a balanced dataset, and designing an improved model. All these goals are implemented with possible improved strategies in the respective fields. To satisfy the first goal, preprocessing steps like feature extraction, hair removal etc are not considered in this work. Rather, the direction of work flow diverted in other aspects of analysing the images. Class imbalance problem is verified whether it really afftects the peroformance of the classifying model or not. From the experiment, the class imbalance problem is observed and avoided by balancing the dataset with SMOTE-TOMEK Links model and target-based augmentation method. Balancing the dataset was not enough to provide improved performance, therefore, an improved classifying model is designed considering the enselmble learning technique with three SVMs. The overall combination of these steps provided a better detection of cancer from dermatoscopic images.

The performance variation using two different types of data and two different types of models is observed in Fig. 6 and 7. The original dataset is imbalanced in classes and cannot train the models effectively. Still, a performance improvement is observed using the stacked ensemble model. The single SVM provided 84%, 82%, and 80.38% training, validation, and testing score, respectively, whereas EnsembleSVM provided slightly higher values, such as 88.72%, 85.76%, and 83.13% % training, validation, and testing score, respectively. The augmented and balanced dataset proved to be a better source for increased performance. The EnsembleSVM model provided the highest values of training, validation, and testing scores as 99.9%, 98%, and 98%, respectively. The confusion matrices provided in Fig. 8 show the performance increment. The false predictions are very low for the proposed EnsembleSVM trained with augmented data of skin dermatoscopic images. From Table 4, it is observed that various preprocessing steps and feature extraction steps have been adopted in the state-of-art methods. Cancer detection is proposed in this work with minimum preprocessing steps and avoidance of any extra feature extraction step. The ultimate effort is to design an appropriate classifier that receives the raw images as input and accurately predicts. The proposed EnsembleSVM model provides higher accuracy with a simple structure and only one preprocessing step, i.e., augmentation for class balancing. The assumptions needed for a real-time system are as follows: the design must make predictions on real-time data, and each time new data will come, the system must be updated with the new data, i.e., the model must continue learning. The proposed

system is a real-time system. This model can be integrated with dermatoscopic image-capturing devices for cancer detection on real-time data.

5. Conclusion

In this work, ensemble learning for a sophisticated machine learning model, i.e., SVM, is adopted to detect skin cancer from dermatoscopic images automatically. The class imbalance problem is verified by considering a single and an ensemble model. It is overcome by using SMOTE-TOMEK Links and a novel data augmentation technique where a target is set for class balancing. The classifying model is developed with three SVMs using the ensemble learning method, and the performance is found to be competitive with the recent deep learning models. The proposed EnsembleSVM model provided a 99.9% training score and a 98% testing score which shows the effectiveness of the proposed model. The effect of class imbalance and augmentation will be verified with other deep-learning models in the future. The weakness of this work is that it is well-fitted with the skin cancer dataset, whereas the performance of the same model is not verified with other cancer data. This issue will be resolved in the future, and a robust model will be designed to work for different types of cancer data.

Statements and Declarations

Funding: The authors did not receive support from any organization for the submitted work.
Conflict of Interest: The authors declare that they have no conflict of interest.
Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.
Informed Consent: This article does not contain patient data.

References

[1] IAFR Cancer. Global Cancer Observatory. Available from: https://gco.iarc.fr/today/online-analysis-table (Accessed on 11th June 2022)

[2] Dildar, M., Akram, S., Irfan, M., Khan, H.U., Ramzan, M., Mahmood, A.R., Alsaiari, S.A., Saeed, A.H.M., Alraddadi, M.O. and Mahnashi, M.H., 2021. Skin cancer detection: a review using deep learning techniques. International journal of environmental research and public health, 18(10), p.5479.

[3] Argenziano, G., Fabbrocini, G., Carli, P., De Giorgi, V., Sammarco, E., & Delfino, M. (1998). Epiluminescence microscopy for diagnosing suspicious melanocytic skin lesions: comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis. Archives of dermatology, 134(12), 1563-1570.

[4] Thanh, D. N., Prasath, V. B., Hieu, L. M., & Hien, N. N. (2020). Melanoma skin cancer detection method based on adaptive principal curvature, color normalization, and feature extraction with the ABCD rule. Journal of Digital Imaging, 33(3), 574–585.

[5] She, Z., Liu, Y., & Damatoa, A. (2007). Combination of features from skin pattern and ABCD analysis for lesion classification. Skin Research and Technology, 13(1), 25-33.

[6] Mohanaiah, P., Sathyanarayana, P., & GuruKumar, L. (2013). Image texture feature extraction using the GLCM approach. International journal of scientific and research publications, 3(5), 1-5.

[7] Ou, X., Pan, W., & Xiao, P. (2014). In vivo skin capacitive imaging analysis by using grey level co-occurrence matrix (GLCM). International Journal of Pharmaceutics, 460(1-2), pp. 28–32.
[8] Abbas, Z., Rehman, M. U., Najam, S., & Rizvi, S. D. (2019, February). An efficient gray-level co-occurrence matrix (GLCM) based approach towards the classification of skin lesions. In 2019 Amity International Conference on Artificial Intelligence (AICAI) (pp. 317–320). IEEE.
[9] Murugan, A., Nair, S. A. H., & Kumar, K. P. (2019). Detection of skin cancer using SVM, random forest, and kNN classifiers. Journal of medical systems, 43(8), 1–9.
[10] Wang, W. (2022). Further results on mean-square exponential Input-to-State stability of stochastic delayed Cohen-Grossberg neural networks. Neural Processing Letters, pp. 1–13.
[11] Huang, C., Liu, B., Yang, H., & Cao, J. (2022). Positive almost periodicity on SICNNs incorporating mixed delays and D operator. Nonlinear Analysis: Modelling and Control, 27(4), 719-739.
[12] Das, A., Mohapatra, S. K., & Mohanty, M. N. (2021, October). A Concatenated Model with Deep Learning Technique for Skin Lesion Detection. In 2021 5th International Conference on Information Systems and Computer Networks (ISCON) (pp. 1–5). IEEE.
[13] Das, A., Mohapatra, S. K., & Mohanty, M. N. (2022). Design of deep ensemble classifier with fuzzy decision method for biomedical image classification. Applied Soft Computing, 115, 108178.
[14] Das, A., Mohanty, M. N., Mallick, P. K., Tiwari, P., Muhammad, K., & Zhu, H. (2021). Breast cancer detection using an ensemble deep learning method. Biomedical Signal Processing and Control, p. 70, 103009.
[15] Hasan, H., Shafri, H. Z., & Habshi, M. (2019, November). A comparison between support vector machine (SVM) and convolutional neural network (CNN) models for hyperspectral image classification. In IOP Conference Series: Earth and Environmental Science (Vol. 357, No. 1, p. 012035). IOP Publishing.
[16] Rastghalam, R., Danyali, H., Helfroush, M. S., Celebi, M. E., & Mokhtari, M. (2021). Skin melanoma detection in microscopic images using HMM-based asymmetric analysis and expectation maximization. IEEE Journal of Biomedical and Health Informatics, 25(9), 3486-3497.
[17] Alfed, N., & Khelifi, F. (2017). Bagged textural and color features for melanoma skin cancer detection in dermatoscopic and standard images. Expert Systems with Applications, 90, 101-110.
[18] Ain, Q. U., Al-Sahaf, H., Xue, B., & Zhang, M. (2020). Generating knowledge-guided discriminative features using genetic programming for melanoma detection. IEEE Transactions on Emerging Topics in Computational Intelligence, 5(4), 554-569.
[19] Tumpa, P. P., & Kabir, M. A. (2021). An artificial neural network-based detection and classification of melanoma skin cancer using hybrid texture features. Sensors International, 2, 100128.
[20] Tan, T. Y., Zhang, L., Neoh, S. C., & Lim, C. P. (2018). Intelligent skin cancer detection using enhanced particle swarm optimization. Knowledge-based Systems, pp. 158, 118–135.
[21] Yu, Z., Nguyen, J., Nguyen, T. D., Kelly, J., Mclean, C., Bonnington, P., ... & Ge, Z. (2021). Early Melanoma Diagnosis with Sequential Dermatoscopic Images. IEEE Transactions on Medical Imaging, 41(3), 633–646.
[22] Usmani, U. A., Watada, J., Jaafar, J., Aziz, I. A., & Roy, A. (2021). A Reinforcement Learning Algorithm for Automated Detection of Skin Lesions. Applied Sciences, 11(20), 9367.
[23] Kwasigroch, A., Grochowski, M., & Mikołajczyk, A. (2020). Neural architecture search for skin lesion classification. IEEE Access, 8, 9061-9071.
[24] Song, L., Lin, J., Wang, Z. J., & Wang, H. (2020). An end-to-end multi-task deep learning framework for skin lesion analysis. IEEE Journal of Biomedical and health informatics, 24(10), 2912-2921.
[25] Adegun, A. A., & Viriri, S. (2019). Deep learning-based system for automatic melanoma detection. IEEE Access, 8, 7160-7172.

[26] Wei, L., Ding, K., & Hu, H. (2020). Automatic skin cancer detection in dermoscopy images based on ensemble lightweight deep learning network. IEEE Access, 8, 99633-99647.
[27] Wang, Y., Louie, D. C., Cai, J., Tchvialeva, L., Lui, H., Wang, Z. J., & Lee, T. K. (2021). Deep learning enhances polarization speckle for in vivo skin cancer detection. Optics & Laser Technology, 140, 107006.
[28] Kumar, K. S., Suganthi, N., Muppidi, S., & Kumar, B. S. (2022). FSPBO-DQN: SeGAN-based segmentation and Fractional Student Psychology Optimization enabled Deep Q Network for skin cancer detection in IoT applications. Artificial Intelligence in Medicine, p. 129, 102299.
[29] Hsiao, Y. P., Chiu, C. W., Lu, C. W., Nguyen, H. T., Tseng, Y. S., Hsieh, S. C., & Wang, H. C. (2021). Identification of Skin Lesions by Using a Single-Step Multiframe Detector. Journal of Clinical Medicine, 10(1), 144.
[30] Courtenay, L.A., González-Aguilera, D., Lagüela, S., Pozo, S.D., Ruiz, C., Barbero-García, I., Román-Curto, C., Cañueto, J., Santos-Durán, C., Cardeñoso-Álvarez, M.E. & Roncero-Riesco, M. (2022). Deep Convolutional Neural Support Vector Machines for Classification Basal Cell Carcinoma Hyperspectral Signatures. Journal of Clinical Medicine, 11(9), p.2315.
[31] Shahin, A. H., Kamal, A., & Elattar, M. A. (2018, December). Deep ensemble learning for skin lesion classification from dermatoscopic images. In 2018 9th Cairo International Biomedical Engineering Conference (CIBEC) (pp. 150–153). IEEE.
[32] Nahata, H., & Singh, S. P. (2020). Deep learning solutions for skin cancer detection and diagnosis. In Machine Learning with Health Care Perspective (pp. 159-182). Springer, Cham.
[33] Bajwa, M.N., Muta, K., Malik, M.I., Siddiqui, S.A., Braun, S.A., Homey, B., Dengel, A. & Ahmed, S. (2020). Computer-aided diagnosis of skin diseases using deep neural networks. Applied Sciences, 10(7), 2488.
[34] Luque, A., Carrasco, A., Martín, A., & de Las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. Pattern Recognition, 91, 216-231.
[35] Mohanty, M. N., & Das, A. (2023). Skin Cancer Detection from Dermatoscopic Images Using Hybrid Fuzzy Ensemble Learning Model. International Journal of Fuzzy Systems, 1-14.
[36] Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. Journal of Big Data, 6(1), 1-54.
[37] Cao, Z., Li, Z., Zhang, J., & Fu, H. (2022). A Homogeneous Stacking Ensemble Learning Model for Fault Diagnosis of Rotating Machinery with Small Samples. IEEE Sensors Journal, 22(9), 8944–8959.
[38] Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset extensively collects multi-source dermatoscopic images of common pigmented skin lesions. Scientific data, 5(1), 1-9.
[39] Batista, G. E., Bazzan, A. L., & Monard, M. C. (2003). Balancing training data for automated annotation of keywords: a case study. Wob, 3, 10-8.
[40] Ben-Hur, A., Ong, C. S., Sonnenburg, S., Schölkopf, B., & Rätsch, G. (2008). Support vector machines and kernels for computational biology. PLoS computational biology, 4(10), e1000173.
[41] Kumar, A., Mohanty, M. N., & Routray, A. (2010, April). Design of support vector machines with time-frequency kernels for classification of EEG signals. In 2010 IEEE Students Technology Symposium (TechSym) (pp. 330–333). IEEE.
[42] Keerthi, S. S. (2002). Efficient tuning of SVM hyperparameters using radius/margin bound and iterative algorithms. IEEE Transactions on Neural Networks, 13(5), 1225-1229.
[43] Géron, A. (2019). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. " O'Reilly Media, Inc.".

[44] Savas, C., & Dovis, F. (2019). The impact of different kernel functions on the performance of scintillation detection based on support vector machines. Sensors, 19(23), 5219.

Abhishek Das is presently working as an Assistant Professor in the Department of Computer Science & Engineering, Centurion University of Technology and Management, Paralakhemundi, Odisha, India. He received his Ph. D. Degree and M. Tech Degree from Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India in 2023 and 2016 respectively. He has completed his B. Tech Degree from Biju Patnaik University of Technology, Odisha in 2013. He is working in the area of biomedical image processing and machine learning. His research interests include machine learning, soft computing, deep learning, swarm and evolutionary computing applications to signal and image processing.

Mihir Narayan Mohanty is presently working as a Professor in the Department of Electronics and Communication Engineering, Institute of Technical Education and Research, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India. He has published over 400 papers in international/national journals and conferences along with approximately 28 years of teaching experience. He is an active member of many professional societies like IAOE, IEEE, IET, EMC & EMI Engineers India, ISCA, ACEEE, IAEng, CSI, and also a Fellow of IETE and IE (I). He has received his M. Tech in Communication System Engineering from the Sambalpur University, Sambalpur, Odisha, and completed his Ph.D. work in Applied Signal Processing. His area of research interests includes applied signal and image processing, biomedical signal processing, microwave communication engineering, system identification, and speech processing.