Tech Science Press

Intelligent Multiclass Skin Cancer Detection Using Convolution Neural Networks

Reham Alabduljabbar[*] and Hala Alshamlan

Department of Information Technology, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia
[*]Corresponding Author: Reham Alabduljabbar. Email: ralabduljabbar@ksu.edu.sa

Abstract: The worldwide mortality rate due to cancer is second only to cardiovascular diseases. The discovery of image processing, latest artificial intelligence techniques, and upcoming algorithms can be used to effectively diagnose and prognose cancer faster and reduce the mortality rate. Efficiently applying these latest techniques has increased the survival chances during recent years. The research community is making significant continuous progress in developing automated tools to assist dermatologists in decision making. The datasets used for the experimentation and analysis are ISBI 2016, ISBI 2017, and HAM 10000. In this work pertained models are used to extract the efficient feature. The pertained models applied are ResNet, InceptionV3, and classical feature extraction techniques. Before that, efficient preprocessing is conducted on dermoscopic images by applying various data augmentation techniques. Further, for classification, convolution neural networks were implemented. To classify dermoscopic images on HAM 1000 Dataset, the maximum attained accuracy is 89.30% for the proposed technique. The other parameters for measuring the performance attained 87.34% (Sen), 86.33% (Pre), 88.44% (F1-S), and 11.30% false-negative rate (FNR). The class with the highest TP rate is 97.6% for Melanoma; whereas, the lowest TP rate was for the Dermatofibroma class. For dataset ISBI2016, the accuracy achieved is 97.0% with the proposed classifier, whereas the other parameters for validation are 96.12% (Sen), 97.01% (Pre), 96.3% (F1-S), and further 3.7% (FNR). For the experiment with the ISBI2017 dataset, Sen, Pre, F1-S, and FNR were 93.9%, 94.9%, 93.9%, and 5.2%, respectively.

Keywords: Convolution neural networks; skin cancer; artificial intelligence; dermoscopy; feature extraction; classification

1 Introduction

Cancer is a disease that occurs due to the unrestricted growth of the irregular cells in the body. These irregular cells have the potential to replicate, dive, and spread in other body parts, such as the lymph and blood. These cells destroy the regular tissues of the body [1]. The worldwide mortality rate due to cancer is second to cardiovascular disease. The International

Agency for Research on Cancer reported the death of approximately 9 million patients because of cancer and 18 million new cases worldwide in 2018 [2]. Cancer growth is due to environmental issues, such as pollution in the air, history of cancer in the family, and lousy lifestyle, including alcohol consumption and addiction to smoking. These are causing harm to the deoxyribonucleic acid (DNA), ultimately leading to cancer. The mortality rate due to cancer could be effectively controlled, but more research is required. The discovery of image processing, latest artificial intelligence techniques, and upcoming algorithms can be used to diagnose and prognose the disease considerably fast and effectively. Applying these latest developments and techniques for this disease has increased survival chances during recent years.

There are various types of cancers, few of them are as follows: (1) Carcinoma is a type of cancer that occurs in the skin and other parts of the body, such as the pancreas, lungs, and breasts. (2) Sarcoma is a type of cancer usually appears in the bone and muscles and even at the connective tissues of the body. (3) Leukemia occurs in the bloodwhich forms tissues like bone marrow and leads to irregular blood cells. (4) Lymphoma occurs in the cells of the immune system. (5) The central nervous system cancer appears in the spinal cord and brain. (6) Melanoma is a kind of skin cancer, which starts in the cell and then forms a pigment in the skin and finally spreads to other organs, as shown in Fig. 1. According to [3], skin cancer is one of the deadly types of cancer that has spread worldwide. Similarly, according to [4] the disease is growing with the report of approximately 10,000 new cases per month globally. The cause of this disease is the exposure of skin to ultraviolet (UV) rays, which are coming from the sun, and these rays cause harm to the DNA of the skin cells. Apart from this, genetic defects are also considered as the vital source of this kind of cancer [5].
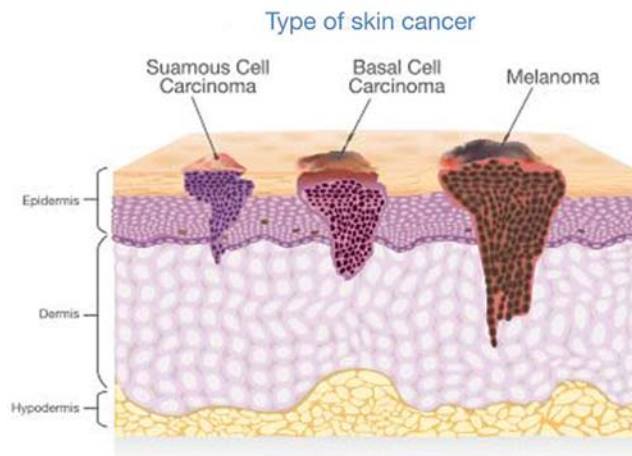
Figure 1: Types of skin cancer

The following are the classes of skin cancer: Melanoma (MEL), Melanocytic nevus (NV), Basal cell carcinoma (BCC), Actinic Keratosis (AK), Benign keratosis lesion, Dermatofibroma (DF), Vascular lesion (VASC), and Squamous cell carcinoma (SCC). Some of them are shown in Fig. 1. The parts of the body that are directly exposed to the sun, such as the head, neck and arms, have more chances of getting the BCC and SCC types of cancers. However, most of these cancers are very common and easy to remedy. Further, MEL is a relatively fast-growing cancer, which spreads rapidly compared to other cancer types. MEL is rare and represents only 5% of all the cancer types, but it is held responsible for more than 70% of the mortality affected

due to skin cancer [6]. Based on this ratio, it is very crucial to handle this type of cancer and classify it at the right time, usually at the early stages, to control the probability of mortality and decrease the mortality rate [7]. The classical method for recognizing MEL requires expert dermatologists to counter the problem of similarity between inter-class and intraclass for the detection of skin lesions. Owing to this issue, the automatic classification of MEL is required to classify the skin lesions with accuracy. This automated system is expected to have enhanced accuracy and enough efficiency to detect the skin lesion. If the detection is done at the early stage, skin cancer could be cured [8]. In the current scenario of skin cancer detection, the clinical experts are performing visual examinations. In the traditional methodology, the analysis is done based on biopsy; testing is done by histopathological study and further dermoscopic assessment [9]. The classification of the skin lesion plays an important role in the early detection and diagnosis of skin cancer. However, to do so, precise expertise is required, which unfortunately is not available with the traditional method of clinical activities. There is a study called dermoscopy, which helps in detecting the skin lesion through skin surface microscopy. This is done to evaluate the pigmented skin lesion. This process is done with the images of the skin. This technique has assisted dermatologists to enhance the accuracy of diagnosis compared to the unaided visual inspection [10,11]. The classification of the skin lesion is based on the following main features: the color features, dermal features, contour features, geometric features, and texture features of the lesions. The full flow of the proposed system is shown in Fig. 2. The classification done by visual examination is very risky as the chances of wrong recognition are relatively high because of the high similarity among different lesion classes [12]. Owing to this difficulty, the classification of the skin lesion is done by deep convolution neural network methodology as an alternative solution for the visual examination. Although the deep-learning technique is far better than the traditional methods, it faces few challenges. Some of those challenges are highly imbalanced classes, a high degree of similarity between the interclass and intraclass, and the inclusion of many artifacts, such as hair, gel bubbles, and ink markers in the dermoscopic images, which make the recognition difficult and challenging.
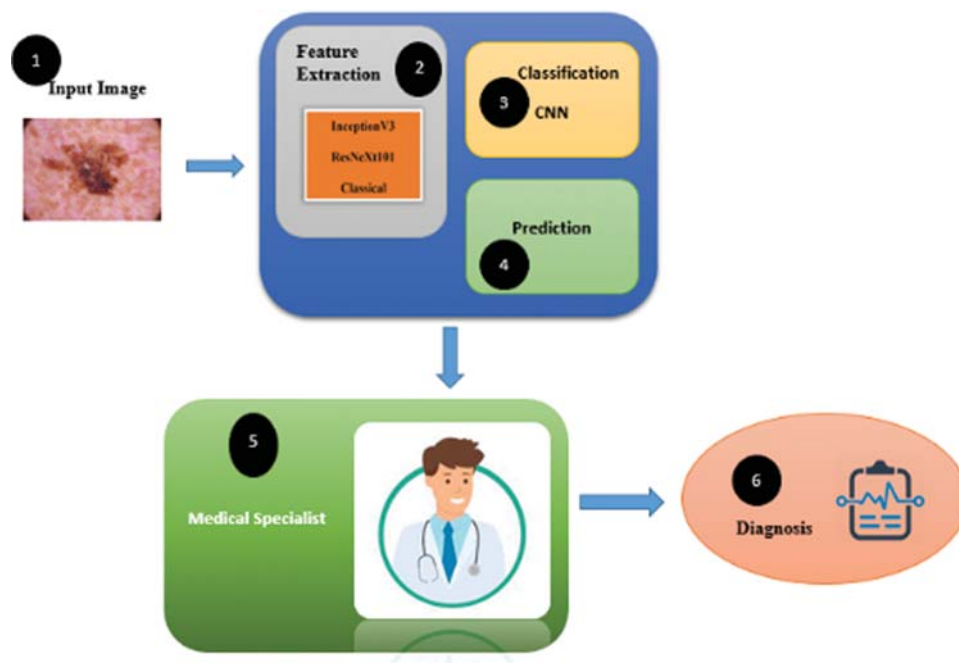
Figure 2: Proposed computer-aided design system

The paper is structured as follows: in Section 2, the related work is discussed. Section 3 gives the details about the three datasets used in this study and also preprocessing and data augmentation steps. Section 4 describes the proposed methodology. Section 5 consists of the Results and Analysis. Section 6 is the conclusion from the study.

2 Related Work

Among many other cancers, skin cancer is a common human malignancy that is typically diagnosed by a visual examination done by experts. The diagnosis starts with a simple clinical screening followed by dermoscopic assessments, such as biopsy and histopathological testing. This traditional diagnosis system has recently been changed to the latest conventional artificial intelligent techniques, which begin with handcrafted feature extraction, then separation training and testing phase followed by actual classification. The classical approach for feature extraction is based on low-level handcrafted features. These features are used to solve classification problems, and can also be utilized for MEL and non-MEL skin lesion classification. However, classical hand-crafted features suffer from poor generalization capability, especially for dermoscopic images and because its biological mechanism is not clearly understood. Based on these issues, it is understood that low-level handcrafted features don't suit a complex disease like skin lesion. Another issue with the handcrafted features is the similarity in visual among inter and intraclass differences. There is no big visible difference, which results in poor performance in classification. In the last decade, there has been a lot of work to classify non-melanocytic and melanocytic skin lesions. In this work [13], the authors have applied a binary mask for computing the principal component analysis. This technique is applied to classify the predictable images for a skin lesion. The accuracy attained by this technique is approximately 82%. To further enhance the accuracy of this system, the authors extracted the features based on cooccurrence and applied these extracted features to classify skin lesions. Applying this technique, they achieved further accuracy of approximately 90%. In another work [14], the authors applied the local binary pattern (LBP) technique to extract the features and then implemented multilayer perceptron, Naïve Bayes, and SVM to classify the skin lesion. Further, with these techniques, they achieved an accuracy of 91.47%. Another feature extraction technique was implemented by [15], which applied morphological high-level intuitive features. This technique could describe the irregularities around the border of the lesion. The authors incorporated this technique to extract low-level features, which would yield a more semantic understanding of the feature, and further, these features would provide the base and foundation for the classification task. These feature extraction techniques produced a classification accuracy of approximately 87.38%. In another work [16], a convolution neural network, along with SVM, was applied for classification. In the preprocessing step of this technique, the texture was extracted for various blocks in the image and analyzed by applying inverse probability along with the LBP technique. The methodology did not get good classification accuracy, which showed up to 71.4%. A computer-aided design (CAD) was implemented by [17] with the help of the (ABCD) rule and feature extraction technique, such as Haralick texture. This system achieved a classification accuracy of approximately 75.1%. Another CAD system was proposed by [18]; it was capable of extracting only the regions with the lesions by applying a texture and color descriptor technique. This technique was applied to non-dermoscopic images. The classification accuracy achieved by this system was 81%. Another work done by [19] applied the ABCD rule for the classification of skin lesions. The authors tried to reduce the noise in the input image. To reduce this noise, they applied a guided filtering technique and further extracted the features of the skin lesion image. This technique achieved a classification accuracy of 79%. In the current scenario for the classification of images, deep learning is considered the most significant, and it has a good

impact in many fields of research, especially in image processing. This technique is a promising tool for extracting the features from the input image by passing it through various models. Deep learning has many layers with neurons, and the layers are filled to extract the fine details from the input image. When the image is passed through the deep-learning technique, it is passed through many layers and finally extracts the relevant features where each layer passes the features to the next layer. This type of extracting features is called Convolutional Neural Network [20]. In another work related to deep learning, the authors in this work [21] used deep networks to detect early-stage skin cancer classification. In this work [22], the authors implemented a deep-learning technique on clinical images to classify skin cancer. They focused on the region-of-interest concept in the preprocessing stage of the experimentation, and this technique enhanced the classification accuracy rate to 81%. In another work related to the CAD systems, the authors [23] introduced a new CAD system by combining supervised and deep-learning algorithms to classify skin images. In the preprocessing stage, they applied the contrast limited adaptive histogram equalization technique to improve the quality of the image, which was used as input for the classification. Afterward, the median filtering technique was applied with the normalized Otsu's segmentation method. These techniques were applied to separate the actually affected lesion from the normal skin. The authors introduced two CAB methods, in which the first system yielded an accuracy of 90.12% using the artificial neural networks, and the second CAD system applied a deep-learning neural network for the classification with an enhanced accuracy of approximately 92.89%. Another CAD system was proposed by [24]; the authors analyzed plain photography. The feature extraction was done using a probabilistic neural network, and further these features were used to decide whether the lesion is melanocytic or non-melanocytic. This work yielded a classification accuracy of approximately 76.2%. Advancing toward DNN, the authors in this work [9] used pertained convolution neural network (CNN) model called Inception v3 to classify the skin lesion. In this work, they increased the input image count using dataset augmentation techniques by rotating the input image to certain angles from 0° to 359° followed by vertical flipping. These augmented images were then applied to the classification, which yielded an accuracy of 71.2%. In another work related to utilizing pre-trained models [25], the authors applied ResNet50. They trained the dataset using the weights of the pre-trained model and achieved the classification accuracy of approximately 69.2% as mean specificity and approximately 76% of sensitivity. In this work [26], a transfer learning technique was applied with a pre-trained model called ResNet152 to classify skin lesions. They did not change the parameters to train the dataset for weight; the original weights were applied during the training process. In this work, three lesions were studied: Seborrheic keratosis, melanoma, and nevus, and the accuracy achieved were 88.2% and 87.63%. The authors in this work [27] extracted ROI and applied them as input images after going through the preprocessing steps for classification. The classification accuracy for the ROI images was 87.2%. In another similar work [28] using ROI, a classification accuracy of 85.5% was achieved with deep residual networks to classify gray-scale images. In this work [29], the authors showed that melanoma is a highly invasive and malignant tumor that can reach the bloodstream easily and cause metastasis in patients [30].

3 Dataset

This research focuses on the dermoscopy images of skin cancer owing to its high impact worldwide [31,32]. In the process of experimentation, we worked on three datasets with varying sizes. The first dataset used was ISBI2016; this dataset consists of 1279 total dermoscopic

images [33]. This dataset was then divided into training and testing sets. In the training set, we took 900 images to train the model, and to test the proposed model, we took 379 images after training the model and tested the model with the respective images. The details are as shown in Tab. 1. The second dataset, on which we trained and tested the model, was ISBI2017. This dataset comprised 2750 dermoscopic images [34]. This dataset was divided into three sets: training, validating, and testing. To train, we selected 2000 images; to validate, 150 images; and to test, 600 images. The details are shown in Tab. 2. Finally, we used HAM10000 dataset [35], which consisted of a total of 10,015 dermoscopic images. This dataset consisted of images with seven different categories. It is a collection of seven different skin lesions, namely: Melanocytic nevi (6705 images), Melanoma (1113 images), Benign keratosis (1099 images), BCC (514 images), AK (327 images), Vascular Lesions (142 images), and Dermatofibroma (115 images). The details are shown in Tab. 3. The sample figures of the HAM 10000 dataset for all the seven classes are shown in Fig. 3. However, these classes were not balanced; therefore, we performed data augmentation as a preprocess for the images before extracting the features and classification. The data augmentation process was done before the training process to balance the input dataset for training and testing. We used 70% of the images for the training, 15% for the validation, and the remaining 15% for the testing. The selection of the images for each stage was done in a random format.

Table 1: ISBI2016 Dataset image details

Dataset	Training	Testing	Total
ISBI2016	900	379	**1279**

Table 2: ISBI2017 Dataset image details

Dataset	Training	Testing	Validation	Total
ISBI2017	2000	600	150	**2750**

Table 3: HAM10000 Dataset image details

Dataset	Classes	Training	Testing	Total
HAM10000	Melanocytic nevi	4690	2015	6705
	Melanoma	770	343	1113
	Benign keratosis	700	399	1099
	Basal cell carcinoma	357	157	514
	Actinic keratosis	227	100	327
	Vascular Lesions	100	42	142
	Dermatofibroma	81	34	115
	Total			**10015**
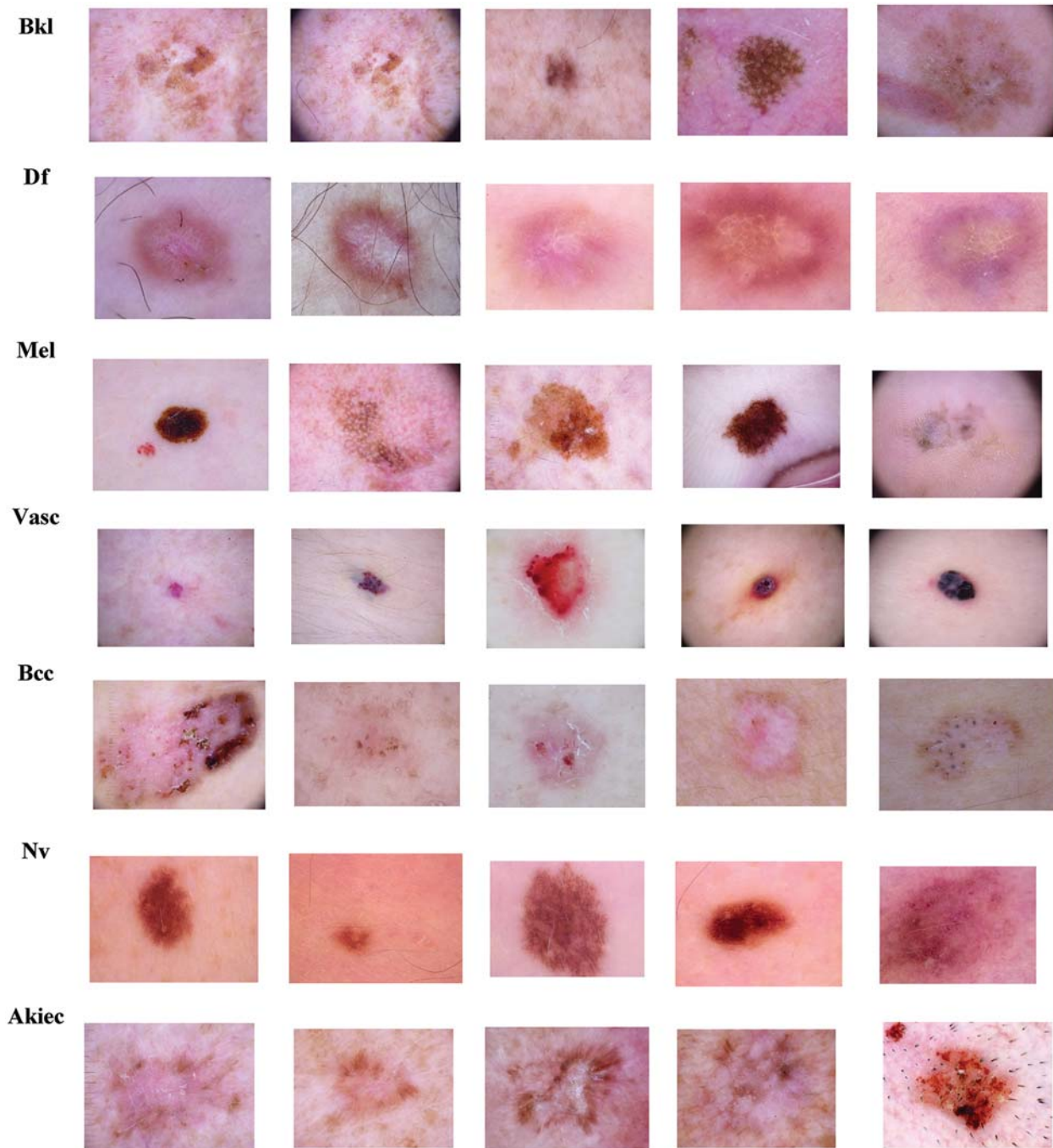
Figure 3: All classes of HAM10000

3.1 Preprocessing

Preprocessing is the stage that is usually done before the data is sent for training and testing. In the preprocessing stage, the input data was reshaped to enhance the classification accuracy and smooth process of training the model. There are various methods for preprocessing the input

data. In this study, for dermoscopic images, we cropped the input images and transformed them into square images, such that the lesion in each image comes to the center of the corresponding image. We also worked on the aspect ratio of the images, such that they were preserved. The input images were rescaled with a resolution of 64×64 pixels. This rescaling was done using nearest interpolation, which preserved the information of the images and decreased the computational cost for processing. The image processing library used for preprocessing the dermoscopic images was OpenCV. During the preprocessing stage, removing extra artifacts, such as hairs, gel bubbles, ink marks, from the dermoscopic images was not required. This is because the proposed intelligent deep-learning technique is capable of handling these kinds of artifacts more intelligently and they will not affect the results. Dataset HAM10000 was partitioned into three. Initially, 70% of the images were partitioned into a training set, and the second partition was allotted for development with 10% of the total images; these developments set images were fine-tuned with the hyper-parameters for the proposed model, and the remaining 20% of the images were allotted for the test set.

3.2 Data Augmentation

Data augmentation techniques were applied to solve the problems of skewed classes, over-fitting issues, scarcity in the number of training images, and underrepresented classes. The preprocessing steps are shown in Fig. 4. These steps were taken to balance the sample sizes for each class during the training process. Hence, the training set was considerably extended to balance the classes. For data augmentation on the training images, the images were rotated with a certain angle, flipped, and translated, as shown in Fig. 4. The images were rotated by $-30°$ to $30°$. For the translation, the images were shifted 12.5% to their left, right, up, and down. Afterward, they were flipped horizontally and vertically. All these operations were applied to the training set only. The testing set images were used without manipulation.
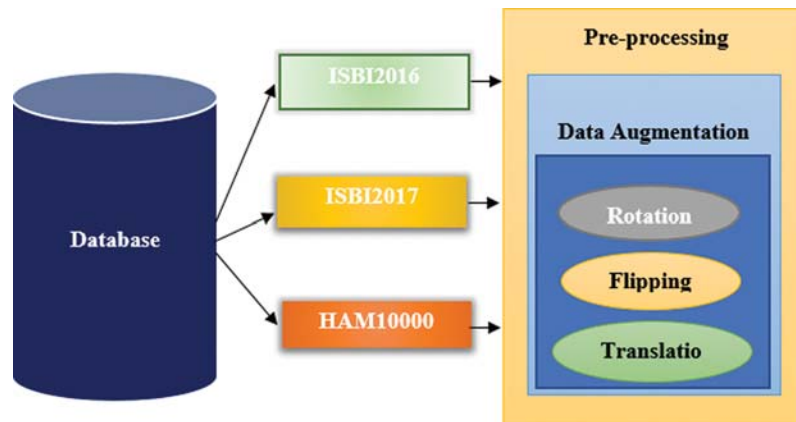

Figure 4: Flow diagram for data preprocessing

4 Proposed Methodology

The CNN was applied in this work since in recent times, this technique is frequently used especially to solve computer vision-related problems. Typically, CNN is a combination of many layers, with each layer doing a particular subtask toward the main classification task. CNN has

subsampling layers called max-pooling or average pooling layer; further, it has a fully connected layer, which is sometimes optional.

The outcome of CNN could be shown using Eq. (1):

$$A_j^1 = f\left(\sum_{i=1}^{M^{l-1}} A_l^{l-1} * w_{ij}^I + b_j^i\right) \tag{1}$$

where feature maps are donated by M^{l-1} for the layer $l-1$, Kernel weight is denoted by w_{ij}^I, the parameter for bias is denoted by b_j^i. Every CNN model has an optimizer that helps in reducing the loss; there are many optimizers available, depending on the requirements for using these optimizers. In this work, we applied stochastic gradient descent and momentum optimizer (SGDM) [36], followed by Adaptive moment estimation (Adam) [37] optimizer. These optimizers were used to control the loss function in the classification and to perform specific fine-tuning to train the model with optimized accuracy. During each iteration, these optimizers helped in updating the weights and biases inside the network to reduce the loss function.

The term momentum was used to avert the oscillations, which were at the steepest descent path. The used SGDM can be expressed as Eq. (2).

$$\theta_{i+1} = \theta_i - \alpha \nabla E_R(\theta_i) + \gamma(\theta - \theta_{i-1}) \tag{2}$$

From the above equation, θ stands for the network parameter vector, and i is the iteration number, and α is the learning rate. α was set as 0.0001 and 0.001 because of the different employed networks. As discussed above, loss functions were applied in our work with ER indicating this loss function, followed by "γ" as the momentum, which was set as 0.9. In this study, we implemented a cross-entropy loss function to optimize the performance as shown in Eq. (3).

$$E_R(\theta) = -\ln\left(\frac{e^{\theta_p}}{\sum_j^c e^{\theta_j}}\right) \tag{3}$$

where θp is the CNN score for the positive class, j represents the iterator number, and C is the number of classes. The minimization of the loss function using Adam is given by Eq. (4).

$$\theta_{i+1} = \theta_i - \frac{\alpha \nabla E(\theta_i)}{\sqrt{V_i} + \varepsilon} \tag{4}$$

where $\beta 2$ is the decay rate, and its value was set to 0.999; another important parameter is "ε", which was maintained at a small value to prevent the zero in the denominator; its value was set to 0.001. To improve the performance of the classification, we worked on the architecture of Xception, InceptionV3, InceptionResNetV2, ResNeXt101, and NASNetLarge. The following are the major customizations carried out to the deep-learning architecture: (a) the dense layer changed with "relu" activation, (b) Softmax and Dropout layers were installed at the bottom of the architecture, and (c) the parameter values are changed. All the above customizations were done to the architecture to enhance the performance of the skin cancer classification. Additional modifications were also done using various fine-tuning CNNs along with ensemble models by implementing stochastic gradient descent (SGD) and Adam optimizers to influence the classification results. All the preprocessing steps are displayed in Fig. 5. The hardware used to implement this methodology was GPU with 6 GB ram and core i7 on windows 10 operating system.
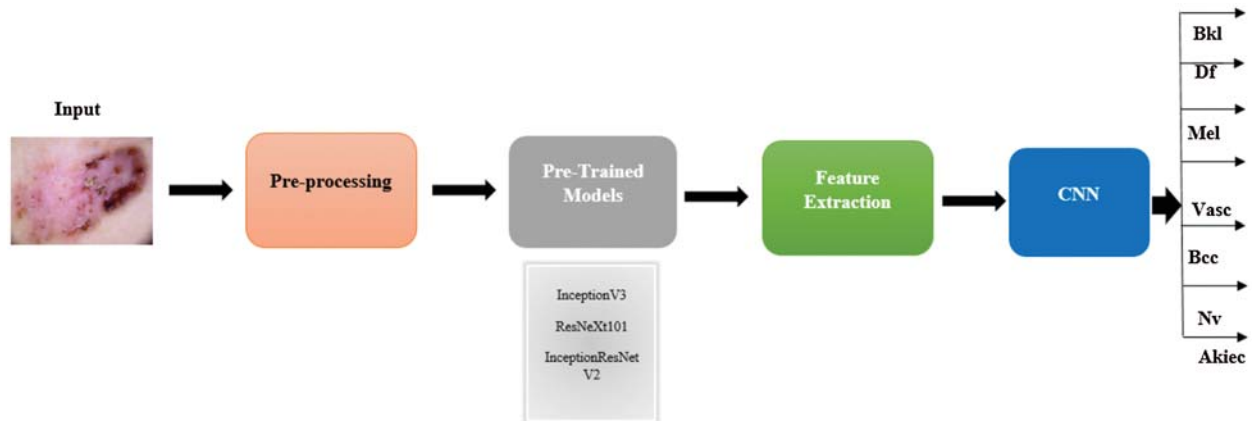
Figure 5: Complete flow of the system along with the convolution neural network and classes

4.1 InceptionV3

InceptionV3 is a network that is very well documented; it is based on inception modules. It consists of a list of convolutions that are arranged parallel to varying kernel sizes to extract the features. The main aim of the InceptionV3 network is to exploit the additional computation and efficiently handling it with suitable factorized convolutions. The InceptionV3 [38] network is considerably capable of handling huge datasets during the training process, which makes it more suitable and popular among researchers as an efficient feature extraction technique. In this work, the dense layer was included along with "relu" activation, and for fine-tuning, dropout and softmax layers, along with seven outputs were installed at the bottom of the architecture of the model on the dataset. Finally, the architecture had 8912 sample images with 30 epochs, a learning rate of 0.0001, and SGD optimizer with a momentum of 0.9.

4.2 ResNeXt101

ResNet is a solution for the accuracy saturation accuracy degradation issues, which comes when the size and depth of the network increases; it proposed an idea that is based on residual connection. There are various models of ResNet, such as ResNet-50, ResNet-101, and ResNet-152. The basic idea for implementing ResNet is the residual learning framework, which provides an easy method to train the relatively deep networks. This capability of ResNet makes it easier to train the model with relatively deep networks, especially during the training phase of the classification cycle. The following are the customization being done to extract the efficient features. The dense layer was included along with "relu" activation, dropout, and softmax layers, along with seven other output modifications to modify ResNet 101 to enhance the accuracy of the classification. The modified ResNeXt101 was then fine-tuned on 8912 images (for 30 epochs) with a learning rate of 0.0001 and SGD optimizer with momentum as 0.9.

4.3 Feature Extraction

In this work, a unique feature extraction methodology was used to extract the feature. These features were extracted using three different pre-trained models, which were discussed in the above section. These pre-trained models were very effective as they reduced the time for developing the deep-learning neural network, especially during the training phase of the process. The prediction accuracy depends on the extracted features; if the feature is very significant, then the results

are also considerably effective and significant. To achieve better accuracy, different features were extracted at different layers. Since each layer affords features, more features can be extracted layer by layer; this would play a significant role in obtaining enhanced classification accuracy. Likewise, other proposed models can also extract more features that would certainly help in identifying skin cancer. A new concept was introduced: transfer learning, which used integrated feature extractors to effectively extract features. The integrated feature extraction concept was applied, in which the model was trained for a particular problem to be solved by applying various fine-tuning options. Based on our problem, it was discovered that using pre-trained networks would be beneficial as the convolution layer, which was relatively close to the input layer, would learn low-level features, such as lines and borders. This capability increased the efficiency of the training to solve other problems. All these pre-trained models and their outputs were integrated with a few of the layers at the end of the architecture. The training process started with the weights of the pre-trained model, and afterward, it was fine-tuned based on the problem to be solved in this work. The weights of the pre-trained model were stationary during the training process since the weights of the pre-trained model were not modified because of the development of a new model for training. All the layers of the CNN plotted their input data to capture the higher-level abstraction. As the features traveled through different layers of the network, more effective features were extracted, which are more informative and effective for classification accuracy. Finally, all the features extracted through a single layer were then deposited in the image for classification. The illustration of the classification process at different layers of the network is shown in Fig. 6, which was performed using the modified Xception model. In [39,40], feature extraction was done by simply training the images using the pre-trained networks followed by the output of the fully connected layers [41]. Some of the other features used by healthcare systems were reported in previous studies [42,43]. However, we hypothesize that the fine-tuning of the pre-trained networks on the relevant dataset can contribute to developing relatively high-quality features, which can boost the performance of the pre-trained models [44].
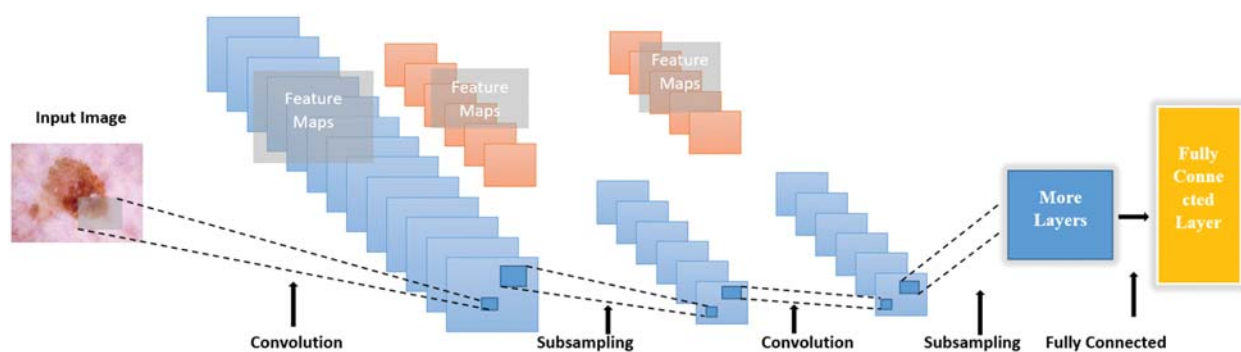
Figure 6: Convolution Neural Network Architecture for the classification of skin cancer

4.4 Performance Matrix

The performance matrix can be evaluated by estimating the predicted image among four subsets: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). TP represents the number of positive cases classified correctly. TN represents the number of negative cases classified correctly. FP, the number of positive cases classified as inaccurate. FN, the number of negative cases classified incorrectly. Accuracy is one of the best measures used to interpret the

performance of models. The accuracy is expressed by using TP, TN, FP, FN as represented by Eq. (5). The other significant performance metric for multi-class classification are precision recall, and F1-score are expressed using Eqs. (6)–(8) respectively.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{5}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{6}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{7}$$

$$F1\text{-}score = 2\left(\frac{Precision * Recall}{Precision + Recall}\right) \tag{8}$$

5 Results and Analysis

The prosed and implemented CAD system for the detection of skin cancer was evaluated with three benchmark datasets, which are described in the datasets section of the paper. The three datasets used to evaluate the proposed CAD system are ISBI2016, ISBI2017, and HAM10000. The classification was done using the proposed methodology, which was discussed in Section 4. The performance of the proposed method to classify the dermoscopic images was evaluated using the following parameters: sensitivity (Sen), precision (Pre), accuracy (Acc), F1-Score (F1-S), and computational time. The proposed method to classify the dermoscopic images was validated on ISBI2016 and ISBI2017 datasets, and the parameters used to evaluate are Pre, Sen, and Mean over coefficient. Apart from these mentioned parameters we also checked the system with overall accuracy, error rate, and time of execution for both the datasets mentioned above. The results obtained with the proposed technique were compared with the standard classifiers, such as KNN, Softmax, Naïve Bayes, and Multi-SVM.

In this section, we will discuss the results obtained with the help of tabular graphs and visual figures. As discussed before in the dataset and preprocessing sections, we divided the whole dataset into training, validating, and testing with the ratio of 70:15:15, respectively. Apart from the above-discussed parameters, we also applied validation techniques such as the 10-fold cross-validation technique. The classification results obtained using the feature extraction technique applied for the proposed methodology for the HAM1000 dataset are discussed in Tab. 4.

Table 4: Classification outcome for HAM10000 Dataset

Approaches	Parameter to assess the performance				
	Sen (%)	Pre (%)	F1-S (%)	Acc (%)	FNR (%)
SoftMax	85.74	86.33	88.44	85.74	11.230
Naïve Bayes	85.23	85.62	84.34	84.56	14.50
M-SVM	83.71	85.24	85.85	82.66	17.92
W-KNN	80.72	83.23	80.31	79.90	16.10
Proposed classifier	**87.34**	**86.33**	**88.44**	**89.30**	**11.30**

The maximum attained accuracy was 89.30% for the proposed technique for the classification of dermoscopic images. Additionally, the other parameters for measuring the performance, which were discussed above, attained 87.34% (Sen), 86.33% (Pre), 88.44% (F1-S), and 11.30% false-negative rate (FNR). The accuracy obtained by other classifiers were 85.74%, 79.90%, 84.56%, and 85.33% for Softmax, KNN, Naïve Bayes, and Multi-SVM, respectively. Furthermore, 18.90% and 11.30% were recorded as the highest and lowest error rates, respectively, in KNN for the proposed method. Further, to validate the results, a confusion matrix is displayed in Tab. 5, which shows the TP rate for each class.

Table 5: Confusion Matrix for HAM10000 Dataset

TSC	Types of skin cancer						
	BKL	Df	Mel	Vasc	BCC	Nv	Akiec
BKL	88.4%	<1%	–	–	<1%	2%	–
Df	<1%	88.1%	<1%	6%	–	<1%	3%
Mel	<1%	–	97.6%	–	2%	–	1%
Vasc	–	2%	2%	93.2%	<1%	<1%	–
BCC	5%	<1%	–	2%	91.2%	–	<1%
Nv	<1%	–	<1%	<1%	–	89.4%	5%
Akiec	–	<1%	–	–	2%	5%	92.3%

The class with the highest TP rate was 97.6% for Mel, whereas the lowest TP rate came for the Df class. The results related to dataset ISBI2016 are shown in Tab. 6. The accuracy achieved for the proposed classifier was 97.0%, whereas the other parameters for validation were 96.12% (Sen), 97.01% (Pre), 96.3 (F1-S), and 3.7% (FNR). In the comparative study with the traditional classifiers, Softmax achieved 95.1%, Naïve Bayes 92.9%, M-SVM 91.9%, and W-KNN 92.1%. As compared to the error rate, the highest was 8.2% for KNN, and the lowest error rate was recorded as 3.6% for the proposed classifier. The accuracy obtained for ISBI2016 with the proposed classifier is shown in Tab. 7 in the form of a confusion matrix. This table shows the true classification for all the classes.

Table 6: Classification results for ISBI2016 Datasets

Approaches	Parameter to assess the performance				
	Sen (%)	Pre (%)	F1-S (%)	Acc (%)	FNR (%)
SoftMax	95.23	94.22	95.34	95.10	4.9
Naïve Bayes	92.33	91.29	91.98	92.97	5.9
M-SVM	91.32	91.98	92.67	91.90	7.1
W-KNN	92.3	92.34	91.66	92.10	8.1
Proposed classifier	**96.12**	**97.01**	**96.34**	**97.87**	**3.7**

Another experiment for the ISBI2017 dataset was conducted and the following are the outcome of the experimentation. The results are shown in Tab. 8; the maximum classification

accuracy attained was 96.9% for the proposed classifier. The experiment was further validated based on other parameters, such as Sen, Pre, F1-S, and FNR, which were 93.9%, 94.9%, 93.9%, and 5.2%, respectively. Similar to the above results, the confusion matrix was applied to check the efficiency of the proposed classifier, which can be seen in Tab. 9. The correctly classified images for both classes were examined. The following are the accuracies obtained with the proposed classifiers: Softmax afforded 93.9%, Naïve Bayes achieved 92.0%, M-SVM gained 90.0%, and KNN reached 91.4%.

Table 7: Confusion matrix for ISBI2016 Dataset

SCT	SCT	
	Benign	Melanoma
Benign	95%	4%
Melanoma	5%	95.5%

Table 8: Classification results for ISBI2017 Datasets

Approaches	Parameter to assess the performance				
	Sen (%)	Pre (%)	F1-S (%)	Acc (%)	FNR (%)
SoftMax	**93.90**	**94.90**	**94.49**	**94.35**	**5.2**
Naïve Bayes	91.21	91.75	91.53	91.69	5.3
M-SVM	86.34	88.34	87.21	87.5	7.6
W-KNN	90.23	90.66	90.51	90.68	5.5
Proposed classifier	**94.90**	**95.30**	**95.89**	**94.89**	**3.5**

Table 9: Confusion matrix for ISBI2017 Dataset

SCT	SCT	
	Benign	Melanoma
Benign	94.4%	5.9%
Melanoma	4.4%	95.5%

6 Conclusions

Cancer is a disease that is due to the unrestricted growth of the irregular cells in the body. These irregular cells have the potential to replicate, dive, and spread in other body parts, such as the lymph and blood. In this work, three datasets were used for experimentation and analysis of skin cancer dermoscopic images. The dataset was applied for preprocessing with efficient data augmentation techniques followed by a few of the pretrained models for the extraction of significant features. These extracted features were applied to the CNN. The following are the results. The maximum attained accuracy is 89.30% for the proposed technique for the classification of dermoscopic images on HAM 1000 Dataset. In contrast, the other parameters for measuring

the performance attained 87.34% (Sen), 86.33% (Pre), 88.44% (F1-S), and 11.30% FNR. The class with the highest TP rate was 97.6% for Mel, whereas the lowest TP rate came for the Df class. In the results related to dataset ISBI2016, the accuracy achieved was 97.0% for the proposed classifier, whereas the other parameters for validation are 96.12% (Sen), 97.01% (Pre), 96.3 (F1-S), and 3.7% (FNR). For the experiment with the ISBI2017 dataset, Sen, Pre, F1-S, and FNR afforded 93.9%, 94.9%, 93.9%, and 5.2%, respectively.

7 Future Scope and Limitations

As a future scope we intend to explore more on the datasets, try to capture real datasets from the hospital staff and get in-depth information about the real disease causes and treatment plans. Another scope is to get comprehensive dataset for training the deep learning algorithm. The dataset preprocessing could also me another future scope as to assist the research community by providing them a unique training dataset exclusively for deep learning techniques which suit to perfectly for learning the features and classifying them further to detect the skin cancer. The work has some limitations as the Convolution Neural Network is applied it needs huge data. To increase the size if the training data we applied data augmentation. But to avoid data augmentation techniques, big dataset is needed.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] H. W. Rogers, M. A. Weinstock, S. R. Feldman and B. M. Coldiron, "Incidence estimate of non-melanoma skin cancer (keratinocyte carcinomas) in the us population 2012," *JAMA Dermatology*, vol. 151, no. 10, pp. 1081–1086, 2015.

[2] A. Stang, K. H. Jöckel and O. Heidinger, "Skin cancer rates in North Rhine-Westphalia, Germany before and after the introduction of the nationwide skin cancer screening program (2000–2015)," *"European Journal of Epidemiology*, vol. 33, no. 3, pp. 303–312, 2018.

[3] K. Chante, A. C. Green, T. Nijsten, M. A. Weinstock, R. P. Dellavalle *et al.,* "The global burden of melanoma: Results from the Global Burden of Disease Study 2015," *British Journal of Dermatology*, vol. 177, no. 1, pp. 134–140, 2017.

[4] B. Harangi, "Skin lesion classification with ensembles of deep convolutional neural networks," *British Journal of Dermatology*, vol. 86, no. 1, pp. 25–32, 2018.

[5] Milroy and J. Mary, "Cancer statistics: Global and National," *Quality Cancer Care*, vol. 2, no. 3, pp. 29–35, 2018.

[6] T. Kanimozhi and A. Murthi, "Computer aided melanoma skin cancer detection using artificial neural network classifier," *Journal of Selected Areas in Microelectronics*, vol. 8, no. 2, pp. 35–42, 2016.

[7] J. F. Anthony, J. T. Johnson, C. D. Sheridan and T. J. Caffrey, "Early detection and treatment of skin cancer," *American Family Physician*, vol. 62, no. 2, pp. 357–368, 2002.

[8] K. M. Hosny, M. A. Kassem and M. M. Foaud, "Classification of skin lesions using transfer learning and augmentation with Alex-net," *PLoS ONE*, vol. 15, no. 5, pp. 1–17, 2019.

[9] E. Andre, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter *et al.,* "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[10] H. Kittler, H. Pehamberger, K. Wolff and M. Binder, "Diagnostic accuracy of dermoscopy," *The Lancet Oncology*, vol. 3, no. 3, pp. 159–165, 2002.

[11] M. E. Vestergaard, P. Macaskill, P. E. Holt and S. W. Menzies, "Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: A meta-analysis of studies performed in a clinical setting," *British Journal of Dermatology*, vol. 159, no. 3, pp. 669–676, 2008.

[12] N. Codella, J. Cai, M. Abedini, R. Garnavi, A. Halpern *et al.,* "Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images," in *Proc. MLMI*, Lima, Peru, pp. 118–126, 2015.

[13] C. W. Yu, A. Huang, C. Y. Yang, C. H. Lee, Y. C. Chen *et al.,* "Computer-aided diagnosis of skin lesions using conventional digital photography: A reliability and feasibility study," *PLoS ONE*, vol. 8, no. 11, pp. 1–9, 2013.

[14] N. Das, A. Pal, S. Mazumder, S. Sarkar, D. Gangopadhyay *et al.,* "An SVM based skin disease identification using Local Binary Patterns," in *Proc. ICACC*, Cochin, India, pp. 208–211, 2013.

[15] R. Amelard, A. Wong and D. A. Clausi, "Extracting morphological high-level intuitive features (HLIF) for enhancing skin lesion classification," in *Proc. EMBS*, San Diego, USA, pp. 1–4, 2012.

[16] E. Karabulut and T. Ibrikci, "Texture analysis of melanoma images for computer-aided diagnosis," in *Proc. ICCSIS 16*, Pattaya, Thailan, pp. 26–29, 2016.

[17] J. A. Almaraz-Damian, V. Ponomaryov and E. Rendon-Gonzalez, "Melanoma CAD based on ABCD rule and haralick texture features," in *Proc. MSMW*, Kharkiv, Ukraine, pp. 21–24, 2016.

[18] I. Giotis, N. Molders, S. Land, M. Biehl, M. F. Jonkman *et al.,* "MED-NODE: A computer-assisted melanoma diagnosis system using non-dermoscopic images," *Expert Systems with Applications*, vol. 42, no. 19, pp. 6578–6585, 2015.

[19] M. H. Jafari, S. Samavi, N. Karimi, S. M. R. Soroushmehr, K. Ward *et al.,* "Automatic detection of melanoma using broad extraction of features from digital images," in *Proc. EMBS*, Florida, USA, pp. 1357–1360, 2016.

[20] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning—An MIT Press Book*. vol. 1, Cambridge: MIT Press, 2016.

[21] A. Esteva, B. Kuprel and S. Thrun, "Deep networks for early stage skin disease and skin cancer classification," *Biomolecules*, vol. 10, no. 1123, pp. 1–15, 2015.

[22] E. N. Esfahani, S. Samavi, N. Karimi, S. M. Soroushmehr, M. H. Jafari *et al.,* "Melanoma detection by analysis of clinical images using convolutional neural network," in *Proc. EMBS*, Florida, USA, pp. 1373–1376, 2016.

[23] J. Premaladha and K. S. Ravichandran, "Novel approaches for diagnosing melanoma skin lesions through supervised and deep learning algorithms," *Journal of Medical Systems*, vol. 40, no. 4, pp. 1–12, 2016.

[24] S. A. Kostopoulos, P. A. Asvestas, I. K. Kalatzis, G. C. Sakellaropoulos, T. H. Sakkis *et al.,* "Adaptable pattern recognition system for discriminating melanocytic nevi from malignant melanomas using plain photography images from different image databases," *International Journal of Medical Informatics*, vol. 105, no. 1, pp. 1–10, 2017.

[25] B. J. Titus, A. Hekler, A. H. Enk, J. Klode, A. Hauschild *et al.,* "A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task," *European Journal of Cancer*, vol. 111, no. 1, pp. 145–154, 2019.

[26] S. S. Han, M. S. Kim, W. Lim, G. H. Park, I. Park *et al.,* "Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm," *Journal of Investigative Dermatology*, vol. 138, no. 7, pp. 1529–1538, 2018.

[27] T. C. Pham, C. M. Luong, M. Visani and V. D. Hoang, "Deep CNN and data augmentation for skin lesion classification," in *Proc. ACIIDS*, Dong Hoi City, Vietnam, pp. 573–582, 2018.

[28] L. Yu, H. Chen, Q. Dou, J. Qin and P. A. Heng, "Automated melanoma recognition in dermoscopy images via very deep residual networks," *IEEE Transactions on Medical Imaging*, vol. 36, no. 4, pp. 994–1004, 2017.

[29] T. Martin, L. Ye, A. Sanders, J. Lane and W. Jiang, "Cancer invasion and metastasis: Molecular and cellular perspective," *Metastatic Cancer Clinical Biological Perspect*, vol. 2, no. 1, pp. 1–12, 2014.

[30] R. Petrilli, J. O. Eloy, F. P. Saggioro, D. L. Chesca, M. V. S. Dias *et al.,* "Skin cancer treatment effectiveness is improved by iontophoresis of EGFR-targeted liposomes containing 5-FU compared with subcutaneous injection," *Journal of Controlled Release*, vol. 283, no. 3, pp. 151–162, 2018.

[31] A. Mahbod, G. Schaefer, I. Ellinger, R. Ecker, A. Pitiot *et al.,* "Fusing fine-tuned deep features for skin lesion classification," *Computerized Medical Imaging and Graphics*, vol. 71, no. 1, pp. 19–29, 2019.

[32] A. Mahbod, G. Schaefer, C. Wang, R. Ecker and I. Ellinge, "Skin lesion classification using hybrid deep neural networks," in *Proc. ICASSP*. Brighton, UK, 1229–1233, 2019.

[33] A. Codella, C. F. Noel, D. Gutman, M. E. Celebi, B. Helba *et al.,* "Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC)," in *Proc. ISBI*, Washington, D.C., USA, pp. 168–172, 2016.

[34] A. Codella, C. F. Noel, D. Gutman, M. E. Celebi, B. Helba *et al.,* "Skin lesion analysis toward melanoma detection: A challenge at the 2017 International symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," in *Proc. ISBI*, Washington, D.C., USA, pp. 150–159, 2018.

[35] P. Tschandl, C. Rosendahl and H. Kittler, "Data descriptor: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data*, vol. 5, no. 1, pp. 1–9, 2018.

[36] A. Hekler, J. S. Utikal, A. H. Enk, A. Hauschild, M. Weichenthal *et al.,* "Superior skin cancer classification by the combination of human and artificial intelligence," *European Journal of Cancer*, vol. 120, no. 3, pp. 114–121, 2019.

[37] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. ICLR*, San Diego, USA, pp. 1–15, 2015.

[38] C. Szegedy, S. Ioffe, V. Vanhoucke and A. A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *Proc. AAAI*, California, USA, pp. 1–12, 2017.

[39] A. R. Lopez, X. G. Nieto, J. Burdick and O. Marques, "Skin lesion classification from dermoscopic images using deep learning techniques," in *Proc. IASTED*, Innsbruck, Austria, pp. 49–54, 2017.

[40] R. B. Oliveira, J. P. Papa, A. S. Pereira and J. M. R. S. Tavares, "Computational methods for pigmented skin lesion classification in images: Review and future trends," *Neural Computing and Applications*, vol. 29, no. 3, pp. 613–636, 2018.

[41] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[42] D. Silva, F. Leno and A. H. R. Costa, "A survey on transfer learning for multiagent reinforcement learning systems," *Journal of Artificial Intelligence Research*, vol. 64, no. 3, pp. 645–703, 2019.

[43] M. A. Khan, M. T. Quasim, N. S. Alghamdi and M. Y. Khan, "A secure framework for authentication encryption using improved ECC for IoT-based medical sensor data," *IEEE Access*, vol. 8, pp. 52018–52027, 2020.

[44] M. T. Quasim, M. A. Khan, Abdullah, M. Meraj, M. Singh *et al.,* "Internet of things for smart healthcare: A hardware perspective," in *Proc. ICOICE*, Yemen, pp. 1–5, 2019.