Multimedia Tools and Applications (2023) 82:23817–23831
https://doi.org/10.1007/s11042-023-14752-z

Detection and classification of dermatoscopic images using segmentation and transfer learning

Hewan Shrestha[1] **· Subash Chandra Bose Jaganathan**[2] **·**
Chandramohan Dhasarathan[3] **· Kannadhasan Suriyan**[4]

Received: 20 March 2022 / Revised: 14 July 2022 / Accepted: 4 February 2023 /
Published online: 21 February 2023

Abstract

With the increase in the number of cases every year, skin cancer stays as one of the most common cancers worldwide. Although dermatologists have been aided with modern research study in detection of cancer, proper treatment of cancer has been a quite challenging task due to the visual appearance of cells. In this experiment, we have studied segmentation and classification algorithms for early detection of cancerous cells, as early detection may increase survival rate of the affected person. HAM10000 dataset has been utilized in this study which has 10,015 different images into seven different classes. Three different types of segmentation algorithms have been studied in this experiment, U-Net, ResUNet and DeeplabV3+. Among all these, DeeplabV3+ appears to outperform the rest of the algorithms giving an overall accuracy of 96.21%, precision and recall of 93.26% and 93% respectively. Few pre-trained models namely ResNet50, ResNet152, SqueezeNet1.1 and DenseNet121 have been utilized for classification of skin cancer. As per the results obtained from the experiment, ResNet152 outperforms the rest of the

✉ Subash Chandra Bose Jaganathan
jsubashme@gmail.com

Hewan Shrestha
shresthahewan12@gmail.com

Chandramohan Dhasarathan
pdchandramohan@gmail.com

Kannadhasan Suriyan
kannadhasan.ece@gmail.com

[1] Department of Computer Science, Saarland University, Saarbrucken, Germany

[2] School of Computing Science and Engineering, VIT Bhopal University, Bhopal-Indore Highway, Kothrikalan, Sehore, Madhya Pradesh 466114, India

[3] Computer Science and Engineering Department, Thapar Institute of Engineering and Technology, Patiala, Punjab, India

[4] Department of Electronics and Communication Engineering, Study World College of Engineering, Coimbatore, Tamilnadu, India

three pre-trained models with an overall training and validation accuracy of 81.75% and 78.51% respectively.

Keywords Melanoma detection · U-net segmentation · Transfer learning · Biomedical image classification

1 Introduction

Dermatoscopy, also termed as dermoscopy, has been a common medical examination these days with the sudden rise of skin cancer patients around the world. It has been found that skin cancer is one of the top 20 most common cancers in the United States and worldwide. As skin cancer is getting common all over the world, it is really crucial to detect the cancerous cell as early as possible so that the mortality can be controlled. Cancer is mainly classified as melanoma or non-melanoma. Melanoma refers to those cells which are harmful where non-melanoma are non-harmful cells. As per the reports available, around 76,380 new cases are found to be affected where about 10,130 deaths were estimated in the United States in 2016 [28, 33]. With early detection and prediction of cancerous cells, it is possible to increase the survival rate by proper diagnosis. Ever since the rise in new cases of melanoma, dermatoscopy has been quite known to the public. However, it is not that easy to detect melanoma in it's early stage which makes it more difficult for dermatologists to cure the disease. Different techniques and procedures have been implemented for detection of the melanoma cells. Techniques ranging from usage of hand-crafted features to segmentation [12, 22] to CNNs, have been worked on for detection and classification of dermatoscopic images to find out the cancerous melanoma and prevent it from getting worse by the early detection procedure. Neural network based segmentation techniques and transfer learning techniques have been implemented in this study.

2 Literature survey

This section provides a brief overview of some latest research work carried on segmentation and classification of skin cancer images.

Iandola et al., used SqueezeNet: AlexNet and Abbas et al., used deep neural network on melanoma detection [1, 10, 29]. Abbasi et. al., used deep learning, Convolution neural network with transfer learning on melanoma detection [2]. Bose et al., Chen et al., Jin et al., used Segmentation techniques and feature extraction techniques [6, 7, 9, 17]. Shrestha et al., used Optimization approach [26]. Zhao et al., used Computer vision and Pattern Recognition. Thurnhofer-Hemsi et al., Arora et al., used Convolutional Networks, Densely Connected Convolutional Networks, Deep Convolutional Features on melanoma detection [4, 30, 35]. Albert et al., Chou et al., Diakogiannis et al., Kassani et al., Simon et al., used a deep learning architectures on melanoma detection [3, 10, 11, 18, 29].Skin lesions are having relatively low contrast and the difference in visual similarity between melanoma and non-melanoma cells which makes early detection a quite tough task. Yu et al. (2017) proposed a deep convolutional neural networks introducing residual learning to prevent the model from being overfitted and also added a fully convolutional residual network (FCRN) for effective lesion segmentation [33]. A two-stage architecture has been approached integrating both FCRN and deep residual networks. Lack of training data makes it difficult for any neural network to perform well.

However, the proposed architecture makes it easier to extract features from the images without the need of worrying about the amount of training data [33].

Detection of melanoma by naked eye is not feasible as it may lead to errors. This is where the usage of artificial intelligence in image processing comes into picture. Image processing techniques help dermatologists to perform proper tests and decide accordingly. Due to the presence of noise, artifacts and various other characteristics in the image, it is really a tough work to differentiate between cancerous and non-cancerous cells [16, 33]. Ichim and Popescu (2020) introduced a unique system which has two levels namely: subjective classifiers and objective classifiers. Objective classifiers are the one that undergo backpropagation algorithm and decides the outcome. Second level classifier, also called as objective classifiers attained an accuracy of 97.5% and F1 score of 97.47% [16]. Cutaneous melanoma is one of the most common type of melanoma which can occur in any part of the skin. One of the important aspect of such type of cell is the Blue-White structure (BWS). Madooei et al. (2019) suggested a multiple instance framework which makes use of probabilisic graphical model. Eventually, the proposed model outperforms in the identification of local features of weakly labeled data [20].

Image recognition in dermatoscopic images have been a difficult task to perform. Yu et al. (2019) have proposed a deep neural network based vector encoding technique for the process of feature extraction from the melanoma infected images. The encoder takes a Fisher Vector (FV) as input and classifies melanoma images using support vector machines [34]. Various internal as well as external factors such as allergies, infections, sun exposure,etc. can cause skin cancer. Due to the fact that dermatoscopic images have visual similarities, melanoma classification into lesions have been a challenging job. Thurnhofer-Hemsi et al. (2021) presented an ensemble learning technique with the combination of DCNN and spaced shifting method for accurate classification of skin lesions. With the spaced shifting method, new input images are obtained which are shifted using displacement vectors [31]. Finally, all the shifted versions are combined together for ensembling. Applying the proposed technique on the popular HAM10000 dataset, accuracy and F1-score have been found to be improving [31, 32].

Every image has a color no matter whichever application we are working on. Most dermatoscopic images are often diagnosed based on the color they possess. When we think of differenting melanoma and non-melanoma based on color, it is observed that melanoma cells mostly appear to be black, red, white and blue-gray. S'aez et al. (2019) presented a color label identification issue that maximizes the posterior probability of a pixel to whichever label it belongs to and it also depends on it's own color value and neighbor's color values. However, it has been observed that the proposed method outperforms state-of-the-art techniques as it attains a F1-score of 0.89, an accuracy of 0.9 and a spearman correlation of 0.831 [24]. Barata et al. (2014) proposed two different techniques for detection of melanoma in dermatoscopic images, one using global methods while the other using local features and bag-of-features classifier. Both the global method and local features having specificity and sensitivity of 80%, 96% and 75%, 100% respectively, perform well on color features [5].

Image degradation could easily occur with the introduction of noise. However, there are many other reasons for any being degraded. When we talk about medical images, image degradation could be resolved by preprocessing the input images. Few traditional image cleaning techniques are still quite useful whereas hair removal technique in the context of skin images require more attention. Although plenty of techniques are available for image preprocessing, hair removal methods are yet to get attention. Kim and Hong (2021) proposed a technique of hair removal by introducing a regenerated set of images using generative adversarial networks. With the proposed method, the important features of the lesions are

Table 1 Comparison of related study

Authors	Techniques applied	Performance metrics
Yu et al. [33]	CNNs with fully Convolutional residual network (FCRN) and deep residual network.	Accuracy, sensitivity, specificity, jaccard index and dice coefficient.
Ichim and Popescu [16]	Subjective classifier with ABCD rule and objective classifier using backpropagation algorithm.	Accuracy and F1-score.
Madooei et al. [20]	Probabilistic graphical model trained using image-level labels.	Accuracy, precision, recall, F1-score and specificity.
Yu et al. [34]	Fisher vector (FV) encoder based technique using support vector machines and chi-squared kernel.	Accuracy, mean average preci- sion and area under the curve.
Thurnhofer-Hemsi et al. [31]	Ensembled deep convolutional neural networks with spaced shifting displacement.	Accuracy, sensitivity, specificity, precision, F1-score and mathew's cor- relation coefficient.
S'aez et al. [24]	Maximizing the posterior proba- bility of a pixel to map with a color value.	Accuracy, precision, recall, F1- score, mathew's correlation co- efficient and spearman's coefficient.
Barata et al. [5]	Global methods and local fea- tures and bag-of-features based classifiers.	Sensitivity and specificity.
Kim and Hong [19]	Generative adversarial networks based image reconstruction by minimizing L1 norm loss.	Accuracy, precision, recall, speci- ficity, F1-score and area under the curve.
Chaturvedi et.al [8]	Fine-tuning over seven classes of HAM10000 dataset.	Accuracy, precision and recall.

kept safe by minimizing the loss associated with the learning [19]. Dermatologists are able to diagnose skin cancer with 62% to 80%. However, there have been many advancements to assist in decision making. Chaturvedi et al. (2020) suggested multi-class skin classification using computer-aided system with ensemble techniques [8]. Comparative study of related work can be seen in Table 1.

3 Materials and methods

This section provides an explanation of the dataset and the algorithms used during the entire research.

3.1 Dataset

For the purpose of detection and classification of dermatoscopic images, one proper dataset with all possible cancer images is required to train different models. Datasets having less number of image samples and those lacking diversity of dermatoscopic images, does not provide good performance with neural networks. Here, the Human Against Machine with 10,000 images, shortly referred as HAM10000, has been used [32]. HAM10000 dataset consists of 10,015 dermatoscopic images which have been collected from wider range of population. This dataset consists of seven different classes namely Actinic Keratosis and Intra-epithelial Carcinoma (akiec), Basal Cell Carcinoma(bcc), Benign Keratosis-like Lesions(bkl), Dermatofibroma (df), Melanoma (mel), Melanocytic Nevi (nv) and Vascular Lesions (vasc). Dataset distribution into training set and testing set can be observed in Table 2.

Table 2 Dataset distribution into train and test set

Class	Train	Test
akiec	260	67
bcc	409	105
bkl	877	222
df	91	24
mel	890	223
nv	5362	1343
vasc	111	31

3.2 Preprocessing stage

Image preprocessing is one of the preliminary work to be performed before doing any sort of segmentation or classification tasks. When raw input images are fed to any sort of deep neural network algorithms, it may not perform well. One of the reasons for those trained models to not perform well is because of the scale range difference amongst the images in the dataset. In the image preprocessing stage, all the images in the dataset are resized to 224 × 224. This is because every class of images within the dataset may not be of same dimension. Resizing of images to same dimesion makes it easier for the neural networks to perform good as all inputs would have same input dimesions. After resizing of the images, normalization has been performed on the entire dataset. Normalization is often perfomed to reduce the scale and make all inputs comparable to each other. Image normalization has been executed with the mean values of (0.485, 0.456, 0.406) and standard deviation values of (0.229, 0.224, 0.225). This normalization is often referred as linear transformation based normalization, which can be defined as:

$$f\ (x) = (x\text{–mean})/\text{std} \tag{1}$$

where, x is the input image, mean represents the mean values and std. refers to the standard deviation values.

3.3 Image augmentation

Whenever a trained model is passed with new data which the model has not trained before, it may be difficult for the model to decide the expected output. This is because images may be in different shape, size and orientation which is completely unknown to the trained model. This issue can be resolved with the augmentation process. Image augmentation is a technique of introducing new images from the exising ones. In this technique, new images are created by slightly changing the original existing image. It is absolutely necessary that sufficient input data is available to train the model in order to achieve better results [27]. However, many datasets currently available lack larger dataset size. Image augmentation fulfills the necessity of large dataset by creating required number of augmented images from the original dataset. Moreover, augmentation process helps in achieving better results and preventing overfitting problem. Here, images have been augmented by cropping, rotating and12 flipping the original set of images. Augmented images from HAM10000 dataset can be seen in Fig. 1.

3.4 Image segmentation

Image segmentation is commonly used digital image processing technique which is focused on partitioning an image into number of parts based on features and properties of pixels of that

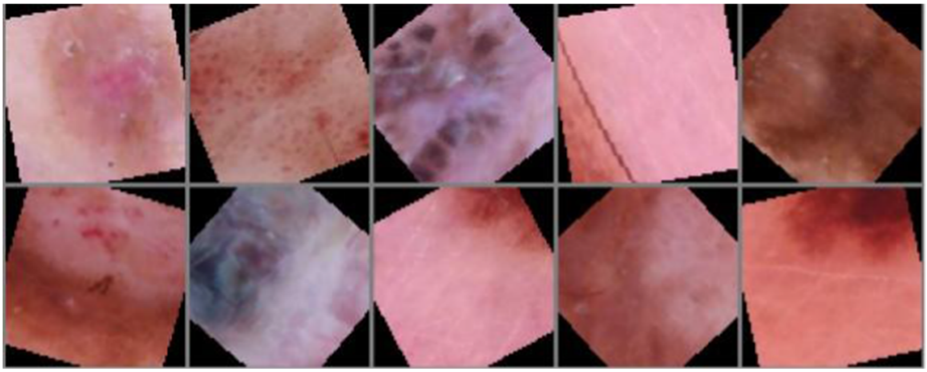

Fig. 1 Augmented images from HAM10000 dataset

image. In segmentation algorithm, the partitioned parts or regions are represented either in the form of contours or masks. As the experiment is focused towards medical diagnosis, some biomedical segmentation algorithms have been used for segmentation of the cancerous region. Three major neural network based segmentation algorithms have been applied on the dataset, namely U-Net, ResUNet and DeeplabV3+.

3.4.1 U-net

U-Net is a FCN based architecture applicable for image segmentation purpose. It is mainly applied in semantic segmentation applications. U-Net was proposed in 2015 by Olaf Ronneburger, Philip Fischer and Thomas Brox at University of Freiburg, Germany [23]. With the traditional convolutional segmentation techniques, it was really time taking process to segment any image. This issue was solved by U-Net as it outperfomed all other segmentation algorithms available during that time. It uses less number of small number of training samples as compared to traditional convolutional approaches. U-Net architecture consists of two paths: Downsampling and Upsampling path as shown in Fig. 2. The Downsampling path in U-Net contains 4 convolution blocks where every block consists of 2 convolution layers, that has 3 × 3 padding followed by ReLU and 2 × 2 max-pooling 29 layer with 2 strides.

Mathematically, ReLU activation function can be denoted as:

$$f(x) = max(0, x) \tag{2}$$

Here, x is any positive input value.

Next to Downsampling path is the Upsampling path, which consists of 4 convolution blocks with 2 convolution layers of 2 × 2 each. With the increase in number of convolution, resolution keeps increasing and the depth keeps decreasing. 3 × 3 filters with ReLU activation function follows the convolution layers. After ReLU, final 1 × 1 convolution layer is added which localizes the regions in the image. Sigmoid activation function is applied to the output image. Mathematically, sigmoid activation function can be given as: $f(x) = \frac{1}{1+e^{-x.}}$

vDeeplabV3+. Batch-normalization (BN) for a channel can be given as:

$$BN(x) = \gamma(x - \mu B/\sigma B) + \beta \tag{4}$$

Here, μ_B and σ_B are the mean and the standard deviation of the batch. γ and β are learned parameters. DeeplabV3+ extracts feature from the backbone network and last blocks of the

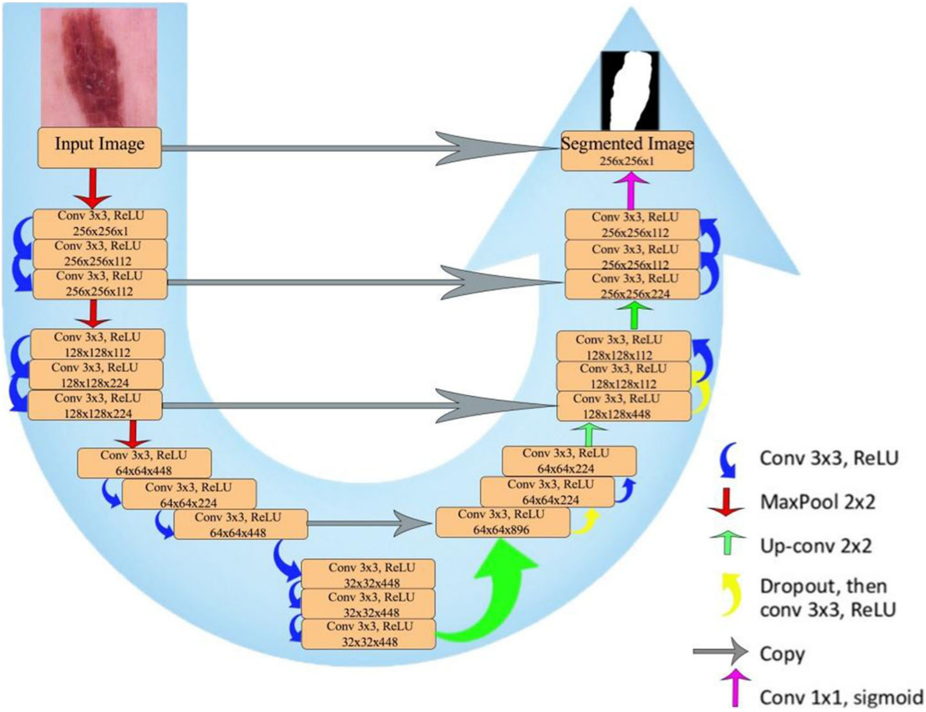

Fig. 2 Architecture of U-Net network

backbone uses atrous convolution. With the usage of atrous convolution, spatial resolution is preserved. As dilation rate is increased through the deeper network, wider filter view is achieved which results in better segmentation results. Some common atrous convolution rates used are 6,12 and 18. After the feature map extraction, 4 parallel atrous convolutions are applied to segment objects at different scales in ASPP network. Average pooling is applied on the last feature map of the backbone. Finally, 1 × 1 convolution is used to get the actual size of the image. This results in creating the final segmented mask for the particular image being convolved. DeeplabV3+ has a total number of 17,869,697 parameters, among which 17,834,913 are trainable and 34,784 are non-trainable parameters as28 shown in Table 3.

3.5 Transfer learning

Transfer learning is a process of making use of already trained model on a new problem which the model has not seen before. It can be used to train deep neural networks with smaller

Table 3 Parameters for segmentation models

Algorithm	Total parameters	Trainable parameters	Non-trainable parameters
UNet	31,055,297	31,043,521	11,776
ResUNet	8,227,393	8,220,993	6400
DeeplabV3+	17,869,697	17,834,913	34,784

amount of data. In transfer learning, initial layers of the neural network remain same with only the later layers being retrained for required computer vision task. Instead of starting the learning process from scratch, with transfer learning, patterns learned earlier to perform a different task. Whenever same feature space is not available for both training and testing, transfer learning comes handy.

There are 2 major elements in transfer learning namely, a Domain and a Task. A Domain contains a feature space x and a marginal probability distribution P (x) where $X = \{x_1, x_2, x_3, x_n\} \in X$. There are 2 components in a Task: a label space Y and a predictive function f (.) The training data is collection of data pairs $\{x_i, y_i\}$ where $x_i \in X$ and $y_i \in Y$ [21]. The source domain can be denoted as $D_S = \{(x_{S1}, y_{S1}), (x_{Sn}, y_{Sn})\}$ where $x_S \in X_S$ is the data point and $y_S \in Y_S$ is the corresponding label. With transfer learning, learning of a new task T_t from the transfer of knowledge from already perfomed task T_S is improved by learning of the function f (.) in the target domain D_t, where $D_S\ f = D_t$ or $T_S\ f = T_t$ [21, 25]. Here, D_S is the source domain, D_t is the target domain and T_S and T_t are the learning tasks. Transfer learning is a technique where we finetune pre-trained models according to the required task to be perfomed. Some of the pre-trained models used are ResNet, SqueezeNet and DenseNet.

3.5.1 ResNet

Whenever we are dealing with convolution neural networks, it is always better to go deep as possible so the model becomes more capable. However, the deeper we move, performance degrades. This problem is solved by introducing ResNets. When the network is too deep, gradients eventually reach to zero, which results in vanishing gradient problem. Gradients are allowed to flow backwards directly through skip connections in ResNets and the vanishing gradient problem gets corrected [13]. There are multiple versions of ResNets based on the number of layers. ResNet versions used here are ResNet50 and ResNet152. ResNet50 has a total of 23.521 M parameters whereas ResNet152 has 58.157 M parameters.

3.5.2 SqueezeNet

SqueezeNet is a CNN architecture that focuses on reducing the number of parameters, without compromising on accuracy. One of the primary objective of squeezenet is to focus on lightweight network. Compared to the state-of-the-art AlexNet, SqueezeNet has 50 times less number of parameters. To reduce the number of parameters, SqueezeNet replaces 3 × 3 filters with 1 × 1 filters. As the filter size is reduced, input channels also need to be reduced. Squeeze layer is introduced to reduce the number of input channels [15]. SqueezeNet1.1 has a total of 421,098 parameters.

3.5.3 DenseNet

DenseNet is one of CNN architectures that focuses on connecting layers directly to each other, which eventually prevents the issue of vanishing gradient. There are a total of $n(n + 1)/2$ direct connections for 'n' layers. The 'n' layer takes input the feature maps obtained from all preceding layers ($i_0, i_1, \ldots i_{n-1}$):

$$i_n = Hn\left(\left[i_0, i_1, \ldots., i_{n-1}\right]\right) \tag{5}$$

Here, [i0, i1, ..., in−1] is the output produced in layers 0, ..., n − 1 [14]. DenseNet uses dense blocks which changes the number of filters keeping the dimensions of feature maps constant. For each layer, Hn is denoted as composite function making use of operations such as batch normalization, a ReLU and a convolution. Whenever the feature map is passed through each dense layer, its size grows due to the addition of 'K' features on top of existing features. The growth rate 'K' monitors the details in each layer of the network. If each function H_n produces k feature maps, the number of input feature maps in the n^{th} layer can be given by Eq. 6.

$$k_n = k0 + k^*(n-1) \tag{6}$$

Here, k_0 is the number of input channels in the network. A bottleneck layer is introduced with 1 × 1 convolution resulting in better efficiency and better computational speed [14]. DenseNet comes in 4 different variants: DenseNet121, DenseNet169, DenseNet201, DenseNet161. In this experiment, DenseNet121 pre-trained model has been applied with a total of 7.2 M parameters.

3.6 Performance metrics

Creating of a model alone is not sufficient for real world applications. It is really important to evaluate whether the trained model performs good with the unknown data. This is where the model evaluation plays a crucial role. There are a number of metrics used to evaluate the performance of a model. However, the evaluation metric depends heavily on the type of task to be performed. Common terms used performance metrics are:

True positives (TP): positive predicted which are actually positive.
False positives (FP): positive predicted which are actually negative.
True negatives (TN): negative predicted which are actually negative.
False negatives (FN): negative predicted which are actually positive.

Metrics used in this experiment are Accuracy, Precision, Recall, Jaccard index and F1 score.

3.6.1 Accuracy

Accuracy is one of the most commonly preferred metric over all other metrics in machine learning tasks. One can easily judge the performance of a model just by looking at its accuracy. In general, accuracy can be defined as the number of correctly predicted data (TP and TN) divided by the total number of data available.

Mathematically, accuracy can be given as:

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

3.6.2 Precision

Precision refers to the number of true positives over the total number of positive predictions (true positives and false positives). In other words, it can be defined as the probability of positive predictions which are actually positive. Precision can be calculated as:

$$\text{precision} = \frac{TP}{TP + FP} \tag{8}$$

3.6.3 Recall

Recall provides the number of true positives over the total number of positive instances (true positives and false negatives). In general, recall measures the fraction of positive values among the total number of positive classes. Recall can be calculated as:

$$\text{recall} = \frac{TP}{TP + FN} \tag{9}$$

3.6.4 Jaccard index

Jaccard index, also termed as Jaccard similarity coefficient, is a metric used to understand the common properties between input datasets. It can be formally defined as the ratio of the size of the intersection to the size of the union of the sample sets. Mathematically, jaccard index or jaccard similarity coefficient can be represented by the formula given below in Eq. 10.

$$\mathrm{J(A, B)} = A \cap B \vee \frac{}{A \cup B \vee} = A \cup B \vee \frac{}{A \vee + B \vee - A \cap B \vee} \tag{10}$$

3.6.5 F1 score

F1 score is the harmonic mean of precision and recall as it combines both of them. This score is just the weighted average of both the precision and recall. The best value for F1 score would be 1 while the worst would be 0. That being said, the higher the F1 score, better is the performance. F1 score can be calculated using the below given formula.

$$\mathrm{F1} = 2 * \frac{precision * recall}{precision + recall} \tag{11}$$

4 Experimentation and results

The implementation of this project has been carried out on Google Colaboratory platform. Colaboratory or Colab for short, is a development tool from Google that is open-sourced and provides free GPU to the public. Colab offers 1 × TeslaK80 GPU which has 2496 CUDA cores and 12GB GDDR5 VRAM. With the use of GPU, parallel processing can be achieved in Google Colab which eventually results in the reduction of training time. For the purpose of model evaluation, multiple metrics have been utilized such as accuracy, precision, recall, jaccard index and F1 score. F1 score combines both the precision and recall into a single value.

4.1 Performance evaluation of segmentation models

For the segmentation part, dataset contains raw images as input and mask as output. These input and output values are preprocessed and augmentation techniques are performed so as to prevent the model from being overfitted and obtain better results. All the segmentation models used are trained for 20 epochs each, with an initial batch size of 16 and a learning rate of 1e-4. With the provided hyperparameters to every segmentation model, it is attained that DeeplabV3+ shows the best overall accuracy of 96.21%, followed by U-Net with accuracy of 94.75%. Predicted mask output with ground truth and input image can be seen in Fig. 3. It is observed that among all 3 segmentation models, DeeplabV3+ has better metrics results, with accuracy of 96.21%, precision of 93.26%, recall of 93%, jaccard index of 86.83% and F1 score of 92.25%. Comparison of all 3 segmentation models can be observed in Table 4 (Figs. 4 and 5).

4.2 Performance evaluation of transfer learning models

With a wide range of pre-trained models available, only few models have been experimented in this study namely, ResNet50, ResNet152, SqueezeNet1.1 and DenseNet121. All the images in the training set undergo normalization and augmentation as part of preprocessing whereas testing set undergo normalization only. All 9 of the pre-trained models are trained for 30 epochs each. Both the training and testing set are normalized with mean and standard deviation. In this study, stochastic gradient descent (SGD) has been used as optimizer with a learning rate of 1e-3. The performance of all 4 pre-trained models can be observed with Table 5.

5 Discussion

With all the past research study conducted over the detection and classification of skin cancer, various results and techniques have been proposed till date. However, the technique of transfer learning which uses pre-trained models and provides better results, have not been explored much with respect to skin cancer. This study fills the gap that has remained unexplored and

Fig. 3 Input image, ground truth and predicted mask using DeeplabV3+

Table 4 Performance comparison of segmentation models

Model	Accuracy	Precision	Recall	Jaccard index	F1 score
U-Net	94.75%	87.66%	92.92%	81.47%	88.71%
ResUNet	91.84%	83.15%	88.45%	73.72%	82.57%
DeeplabV3+	96.21%	93.26%	93%	92.25%	86.83%

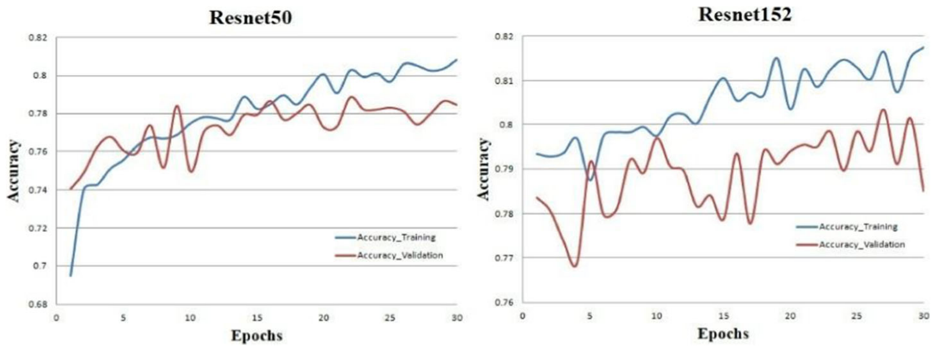

Fig. 4 Accuracy plots for ResNet50 and ResNet152

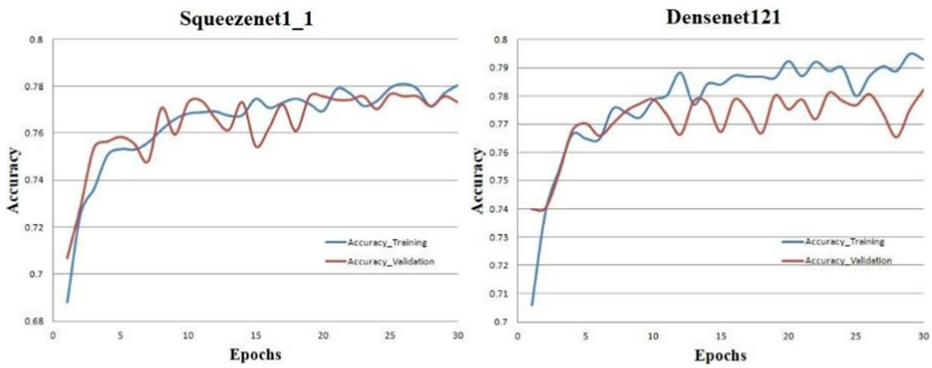

Fig. 5 Accuracy plots for SqueezeNet1_1 and DenseNet121

Table 5 Performance comparision of transfer learning models

Pre-trained Model	Training accuracy	Validation accuracy	Training loss	Validation loss
ResNet50	80.80%	78.46%	51.88%	56.94%
ResNet152	81.75%	78.51%	48.71%	58.53%
SqueezeNet1.1	78.05%	77.32%	59.22%	62.27%
DenseNet121	79.29%	78.21%	56.19%	63.91%

allows researchers to study the field even more. Also, neural network based segmentation techniques have been more helpful for semantic segmentation tasks. Using modern segmentation algorithms, it's even more effective for bio-medical segmentation experiments. Moreover, further advancements can be done over this study. As Generative Adversarial Networks (GANs) are getting more popular, medical image segmentation and classification could be much more reliable and efficient with the usage of GANs.

6 Conclusion

Skin cancer has been one of the deadliest diseases these days and a wide aspect of research is focused towards its early detection. Early detection of cancerous cells is not an easy task to follow. However, few neural network based segmentation algorithms have been implemented in this research study. DeeplabV3+ outperforms the rest of the algorithms with an overall accuracy of 96.21% having precision and recall of 93.26% and 93% respectively. Further upgrades to this study would result in much more effective segmentation and classification. GAN based segmentation could result much better in detecting the cells.

Data availability The data that support the findings of this study are openly available in Harvard Dataverse at https://doi.org/10.7910/DVN/DBW86T [32].

Declarations

Conflict of interest No potential competing interest was reported by the authors.

References

1. Abbas Q, Celebi ME (2019) DermoDeep-A classification of melanoma-nevus skin lesions using multi-feature fusion of visual features and deep neural network. Multimed Tools Appl 78:23559–23580. https://doi.org/10.1007/s11042-019-7652-y
2. Abbasi AA, Hussain L, Awan IA, Abbasi I, Majid A et al (2020) Detecting prostate cancer using deep learning convolution neural network with transfer learning approach. Cogn Neurodyn 14(4):523–533. https://doi.org/10.1007/s11571-020-1609587-5
3. Albert BA (2020) Deep Learning From Limited Training Data: Novel Segmentation and Ensemble Algorithms Applied to Automatic Melanoma Diagnosis. IEEE Access 8:31254–31269. https://doi.org/10.1109/AC-25CESS.2020.2973188
4. Arora R, Raman B, Nayyar K, Awasthi R (2021) Automated skin lesion segmentation using attention-based deep convolutional neural network, biomedical signal processing and control. Volume 65:102358, ISSN 1746-8094. https://doi.org/10.1016/j.bspc.2020.102358
5. Barata C, Ruela M, Francisco M, Mendonça T, Marques JS (2014) Two Systems for the Detection of Melanomas in Dermoscopy Images Using Texture and Color Features. IEEE Syst J 8(3):965–979. https://doi.org/10.1109/JSYST.2013.2271540
6. Bose SC, Kumar K, Karnan M (2012) Detection of microcalcification in mammograms using soft computing techniques. Eur J Sci Res 86(1):103–122
7. Bose SC, Veerasamy M, Mubarakali A, Marina N, Hadzieva E (2020) Analysis of Feature Extraction Algorithm Using Two Dimensional Discrete Wavelet Transforms in Mammograms to Detect Microcalcifications. In: Smys S., Tavares J., Balas V., Iliyasu A. (eds) Computational Vision and Bio-Inspired Computing. ICCVBIC 2019. Advances in intelligent systems and computing, vol 1108. Springer, Cham https://doi.org/10.1007/978-3-030-37218-7_4.
8. Chaturvedi SS, Tembhurne JV, Diwan T (2020) A multi-class skin Cancer classification using deep convolutional neural networks. Multimed Tools Appl 79:28477–28498. https://doi.org/10.1007/s11042-020-09388-2

9. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL (2018) DeepLab: semantic image segmentation with deep convolutional nets, Atrous convolution, and fully connected CRFs. IEEE Trans Pattern Anal Mach Intell 40(4):834–848. https://doi.org/10.1109/TPAMI.2017.2699184
10. Chou HY, Huang SL, Tjiu JW, Chen HH (2021) Dermal epidermal junction detection for full-field optical coherence tomography data of human skin by deep learning. Comput Med Imaging Graph 87:101833. https://doi.org/10.1016/j.compmedimag.2020.101833
11. Diakogiannis FI, Waldner F, Caccetta P, Wu C (2020) ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. ISPRS J Photogramm Remote Sens 162:94–114, ISSN 0924-2716. https://doi.org/10.1016/j.isprsjprs.2020.01.013
12. Dobbs TD, Ottenhof M, Gibson JAG, Whitaker IS, Hutchings HA (2021) The patient-reported outcome measures in skin Cancer reconstruction (PROMISCR) study: Anglicisation and initial validation of the FACE-Q skin Cancer module in a UK cohort. J Plast Reconstr Aesthet Surg 74(3):615–624. https://doi.org/10.1016/j.bjps.2020.09.005
13. He K, Zhang X, Ren S, Sun J (2016) Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 https://doi.org/10.1109/CVPR.2016.90
14. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely Connected Convolutional Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2261–2269, https://doi.org/10.1109/CVPR.2017.243.
15. Iandola, FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K (2016) SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5MB model size, arXiv:1602.07360
16. Ichim L, Popescu D (2020) Melanoma Detection Using an Objective System Based on Multiple Connected Neural Networks. IEEE Access 8:179189–179202. https://doi.org/10.1109/ACCESS.2020.3028248
17. Jin Q, Cui H, Sun C, Meng Z, Ran S (2021) Cascade knowledge diffusion network for skin lesion diagnosis and segmentation. Appl Soft Comput 99:106881. https://doi.org/10.1016/j.asoc.2020.106881
18. Kassani SH, Kassani PH (2019) A comparative study of deep learning architectures on melanoma detection. Tissue Cell 58:76–83
19. Kim D, Hong B-W (2021) Unsupervised Feature Elimination via Generative Adversarial Networks: Application to Hair Removal in Melanoma Classification. IEEE Access 9:42610–42620. https://doi.org/10.1109/ACCESS.2021.3065701
20. Madooei A, Drew MS, Hajimirsadeghi H (2019) Learning to Detect Blue–White Structures in Dermoscopy Images with Weak Supervision. IEEE J Biomed Health Inform 23(2):779–786. https://doi.org/10.1109/JBHI.2018.2835405
21. Pan SJ, Yang Q (2010) A Survey on Transfer Learning. IEEE Trans Knowledge Data Eng 22(10):1345–1359. https://doi.org/10.1109/TKDE.2009.191
22. Putra TA, Rufaida SI, Leu J-S (2020) Enhanced Skin Condition Prediction Through Machine Learning Using Dynamic Training and Testing Augmentation. IEEE Access 8:40536–40546. https://doi.org/10.1109/ACCESS.2020.2976045
23. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: Navab N., Hornegger J., Wells W., Frangi a. (eds) medical image computing and computer-assisted intervention – MICCAI 2015. MICCAI 2015. Lecture notes in computer science, vol 9351. Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4 28
24. S'aez A, Acha B, Serrano A, Serrano C (2019) Statistical Detection of Colors in Dermoscopic Images With a Texton-Based Estimation of Probabilities. IEEE J Biomed Health Inform 23(2):560–569. https://doi.org/10.1109/JBHI.2018.2823499
25. Shell J, Coupland S (2015) Fuzzy transfer learning: methodology and application. Inform Sci 293:59–79, ISSN 0020-0255. https://doi.org/10.1016/j.ins.2014.09.004
26. Shrestha H, Dhasarathan C, Munisamy S, Jayavel A (2020) Natural language processing based sentimental analysis of Hindi (SAH) script an optimization approach. Int J Speech Technol 23:757–766. https://doi.org/10.1007/s10772-020-09730-x
27. Shrestha H, Dhasarathan C, Kumar M, Nidhya R, Shankar A, Kumar M (2022). A deep learning based convolution neural network-DCNN approach to detect brain tumor. In: Gupta, G., Wang, L., Yadav, a., Rana, P., Wang, Z. (eds) proceedings of academia-industry consortium for data science. Advances in intelligent systems and computing, vol 1411. Springer, Singapore. https://doi.org/10.1007/978-981-16-6887-6_11
28. Siegel RL, Miller KD, Jemal A (2015) "Cancer statistics, 2016," CA. Cancer J. Clinicians
29. Simon M (2020) Thomas, James G. Lefevre, Glenn Baxter, Nicholas A. Hamilton, Interpretable Deep Learning Systems for Multi-Class Segmentation and Classification of Non-Melanoma Skin Cancer, Medical Image Analysis https://doi.org/10.1016/j.media.2020.101915

30. Thurnhofer-Hemsi K, Dom'ınguez E (2021) A Convolutional neural network framework for accurate skin cancer detection. Neural Process Lett 53:3073–3093. https://doi.org/10.1007/s11063-020-10364-y
31. Thurnhofer-Hemsi K, Lopez-Rubio E, Dom'ınguez E, Elizondo DA (2021) Skin Lesion Classification by Ensembles of Deep Convolutional Networks and Regularly Spaced Shifting. IEEE Access 9:112193–112205. https://doi.org/10.1109/ACCESS.2021.3103410
32. Tschandl P (2018) The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions", https://doi.org/10.7910/DVN/DBW86T, Harvard Dataverse, V3, UNF: 6:/APKSsDGVDhwPBWzsStU5A==
33. Yu L, Chen H, Dou Q, Qin J, Heng P (2017) Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks. IEEE Trans Med Imaging 36(4):994–1004. https://doi.org/10.1109/TMI.2016.2642839
34. Yu Z et al (2019) Melanoma Recognition in Dermoscopy Images via Aggregated Deep Convolutional Features. IEEE Trans Biomed Eng 66(4):1006–1016. https://doi.org/10.1109/TBME.2018.2866166
35. Zhao H, Shi J, Qi X, Wang X, Jia J (2017) Pyramid Scene Parsing Network. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6230–6239. https://doi.org/10.1109/CVPR.2017.660