Article

An Enhanced Transfer Learning Based Classification for Diagnosis of Skin Cancer

Vatsala Anand [1], Sheifali Gupta [1], Ayman Altameem [2], Soumya Ranjan Nayak [3], Ramesh Chandra Poonia [4] and Abdul Khader Jilani Saudagar [5],*

1 Institute of Engineering and Technology, Chitkara University, Rajpura 140401, Punjab, India; vatsala.anand@chitkara.edu.in (V.A.); sheifali.gupta@chitkara.edu.in (S.G.)
2 Department of Computer Science and Engineering, College of Applied Studies and Community Services, King Saud University, Riyadh 11533, Saudi Arabia; aaltameem@ksu.edu.sa
3 Amity School of Engineering and Technology, Amity University Uttar Pradesh, Noida 201301, Uttar Pradesh, India; nayak.soumya17@gmail.com
4 Department of Computer Science, CHRIST (Deemed to be University), Bangalore 560029, Karnataka, India; rameshcpoonia@gmail.com
5 Information Systems Department, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11432, Saudi Arabia
* Correspondence: aksaudagar@imamu.edu.sa

Abstract: Skin cancer is the most commonly diagnosed and reported malignancy worldwide. To reduce the death rate from cancer, it is essential to diagnose skin cancer at a benign stage as soon as possible. To save lives, an automated system that can detect skin cancer in its earliest stages is necessary. For the diagnosis of skin cancer, various researchers have performed tasks using deep learning and transfer learning models. However, the existing literature is limited in terms of its accuracy and its troublesome and time-consuming process. As a result, it is critical to design an automatic system that can deliver a fast judgment and considerably reduce mistakes in diagnosis. In this work, a deep learning-based model has been designed for the identification of skin cancer at benign and malignant stages using the concept of transfer learning approach. For this, a pre-trained VGG16 model is improved by adding one flatten layer, two dense layers with activation function (LeakyReLU) and another dense layer with activation function (sigmoid) to enhance the accuracy of this model. This proposed model is evaluated on a dataset obtained from Kaggle. The techniques of data augmentation are applied in order to enhance the random-ness among the input dataset for model stability. The proposed model has been validated by considering several useful hyper parameters such as different batch sizes of 8, 16, 32, 64, and 128; different epochs and optimizers. The proposed model is working best with an overall accuracy of 89.09% on 128 batch size with the Adam optimizer and 10 epochs and outperforms state-of-the-art techniques. This model will help dermatologists in the early diagnosis of skin cancers.

Keywords: skin cancer; Kaggle; convolutional neural network; benign; malignant; data augmentation; VGG16; classification

1. Introduction

Cells are the basic building blocks in the human body and are used for building tissues. The skin acts as the outer layer of human beings and shields the human body against infections and harmful radiation. The three different layers of skin are the inner-most dermis, middle hypodermis, and outer epidermis. Skin cancer is the most frequently diagnosed cancer, and the occurrence of skin cancer is growing. Skin cancer is caused by the unusual development of cells. People having fair skin are prone to skin cancer. The two types of skin cancer are malignant and benign [1]. Cancerous cells that grow without any control leads to malignant tumors. Metastasis is a process in which the

cancer cells travel through the lymph nodes or circulation and extend to further parts of the human body. According to the report of the American Cancer Society (ACS), in 2018, 13,460 skin cancer death cases were noted [2,3]. Throughout the years, detecting melanomas via image analysis has shown progression. Most studies of melanomas were based on machine learning algorithms. However, deep learning algorithms have helped in skin lesion classification studies [4].

Haenssle et al. [5] utilized a CNN for classification of dermatoscopy melanocytic images and obtained a value of sensitivity and a specificity as 86.6%. Dorj et al. [6] used an approach of ECOC SVM with an AlexNet pre-trained model for multi-class skin disease classification. They obtained an accuracy of 95.1%. Han et al. [7] used a deep convolutional neural network on 12 different skin diseases. They obtained an accuracy of 96.0%. Khan et al. [8] used various pre-trained architectures like VGG16, DenseNet169, DenseNet161, and ResNet50. In this, they pushed the boundary of neural networks by using low resolution pixels such as 80×80, 64×64, and 32×32. They achieved the highest performance values of 80.46%, 78.56%, and 74.15% for 80×80, 64×64, and 32×32 pixels, respectively. Mohakud et al. [9] has proposed an encoder decoder network for segmentation of image. The authors obtained the value of the Jaccard coefficient as 96.41% and 86.85% respectively, and the Dice coefficient as 98.48% and 87.23%, accuracy as 98.32% and 95.25% respectively for ISIC 2016 and ISIC 2017 dataset.

Agrahari et al. [10] used a pre-trained MobileNet model for building the model and worked using HAM10000 dataset. They obtained categorical accuracy as high as 80.81%. Although Chaturvedi et al. [11] had achieved an overall accuracy of 83.1% using MobileNet architecture with HAM10000 dataset. Hosny et al. [12] used AlexNet model by replacing the last layer by softmax for classification of three skin diseases. They worked on Ph2 dataset and obtained values of accuracy, sensitivity, specificity, and precision as 98.61%, 98.33%, 98.93%, and 97.73%, respectively. Abdar et al. [13] used three uncertainty quantification methods. The accuracy obtained by the model was 88.95%. Fujisawa et al. [14] used 4867 clinical images including benign and malignant conditions. The overall accuracy of the model was 76.5% with sensitivity of 96.3% and specificity of 89.5%. Garcia Arroyo et al. [15] presented an algorithm based on machine learning and used 875 dermoscopic images. The images were collected from the Interactive Atlas of Dermoscopy dataset. Total achieved accuracy was 88.00%, sensitivity was 83.44% and specificity was 90.71%. Iyatomi et al. [16] took 213 dermoscopic images and used a linear classifier. They achieved specificity of 95.90%, and the value of the area under the curve was 0.993. In 2018, Chatterjee et al. [17] used classifiers and took 4094 skin cancer images and accomplished an accuracy of 98.28% and a sensitivity of 97.63%. In 2019, these authors had taken dermoscopic images from the internet and performed GLCM and FRTA feature extraction. They attained 97.54% accuracy [18]. Gonzalez et al. [19] applied the DermaKNet technique on a total of 2750 skin cancer images and achieved an area under the curve value of 91.7%. In 2018, Ka-wahara et al. [20] applied Multi-task multi-modal neural nets on 1011 dermoscopic images. The architecture was able to localize discriminate information and also produce feature vectors.

Koohbanani et al. [21] in 2018 used transfer learning based model and used a total of 2594 dermoscopic images from the internet. This framework incorporates a variant of UNet architecture. Filali et al. [22] presented a network based on CNN by using 1000 images from the internet and realized 93.50% accuracy. Kadampur et al. [23] applied a technique based on machine learning with 104 images of skin from the internet. They realized 86.00% value of sensitivity and 73.00% specificity. For increasing deep learning performance for melanoma screening, Menegola et al. [24] presented a transfer method. A pre-trained model for detecting Diabetic Retinopathy was proposed in this study which was based on Kaggle Challenge [25].

The deep learning revolution has played a great role, with the suggestion of better architectures of the convolutional neural network [26–31]. In the following paper, a model is presented to classify skin cancer with the help of dermoscopy images. The presented

model is evaluated on ten and twenty epochs using the Adam optimizer and batch size values of 8, 16, 32, 64, and 128. The proposed model has presented favorable results that will work as another estimation tool for dermatologists. The major contributions of this study include:

1. A paradigm based on transfer learning has been presented using VGG16 architecture for classification of skin cancer into benign and malignant [32];
2. The VGG16 model has been improved by the addition of one flatten layer, two dense layers with activation function (LeakyReLU) and another dense layer with activation function (sigmoid) to improve the accuracy of the model;
3. The data augmentation techniques have been performed in the pre-processing stage for increasing randomness and dataset count in order to provide stability to the proposed model;
4. The efficacy of the proposed model is achieved by analyzing various hyper parameters such as batch size, epochs, and optimizer.

The rest of the paper is arranged as follows: Section 2 shows the proposed framework model followed by results and discussions in Section 3 and conclusion in Section 4.

2. Proposed Framework Model

A paradigm based on transfer learning has been improved and changed for classification of skin cancer into benign and malignant class. The training and testing of the model is performed on the Kaggle dataset [25] that consists of 3297 skin cancer images. The block diagram of the proposed framework model is shown in Figure 1.
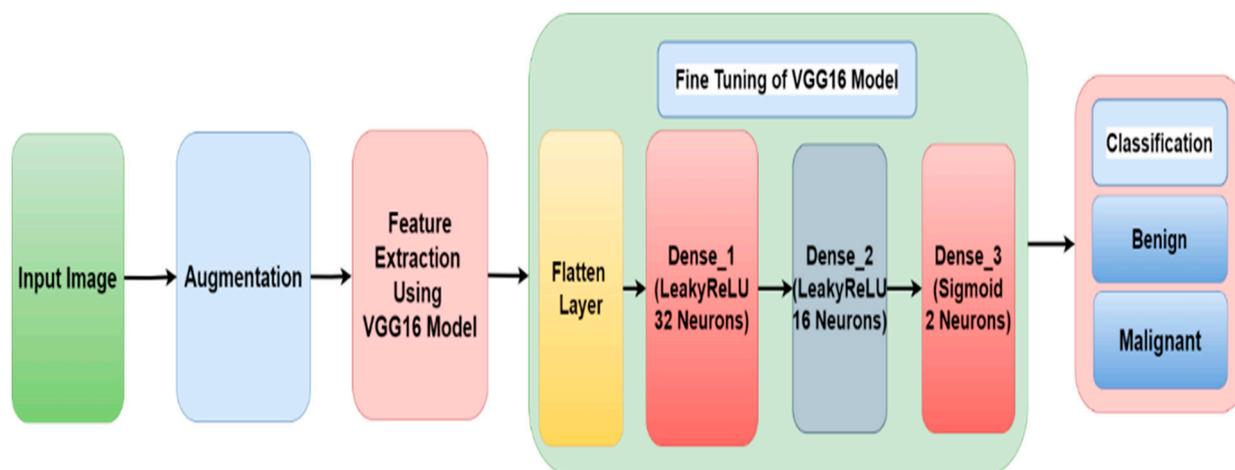

Figure 1. Proposed Framework Model.

2.1. Input Dataset

The database which is used in the study consists of 3297 skin cancer images that are collected from the Kaggle database [25]. It is comprised of RGB images of 1800 benign and 1497 malignant images of dimensions ($224 \times 224 \times 3$) pixels. Figure 2 shows a sample of benign and malignant skin cancer images from the database.

Table 1 shows the dataset description in which the number of training images, testing images and validation images are shown for both skin cancer classes. Total images in the dataset are 3297 out of which 1800 are benign and 1497 are malignant. The dataset is split into testing, training and validation. For the testing purpose, almost 10% of benign and malignant images are used. From the remaining images, 5% of the images are used for validation purposes. The remaining dataset is used for training the model.
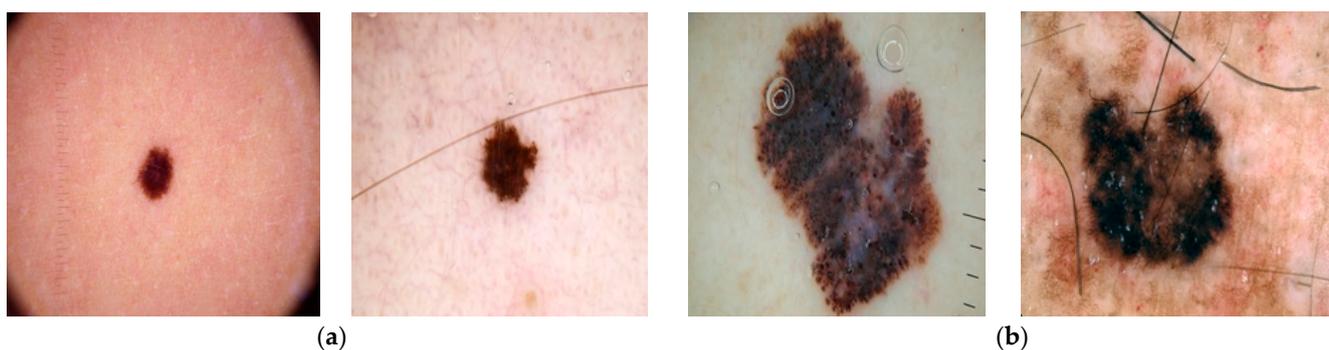
Figure 2. (**a**) Benign, (**b**) Malignant.

Table 1. Dataset Description.

	Total	Training	Test	Validation
Benign	1800	1534	186	80
Malignant	1497	1284	144	69
Total	3297	2818	330	149

2.2. Data Augmentation

A huge quantity of dataset is essential to attain the best accuracy in the DL. The augmentation of data is done with various transformation techniques [33–35] like rotation, flipping, and brightening in sequence as shown in Figure 3. For this, the input image is rotated 90 degrees in a clockwise direction. After that, the rotated image is flipped horizontally as well as vertically. At the end, the brightness level of the flipped image is changed by 0.8. The augmentation process is applied only on training images to train the model more precisely. In this way, the number of training images is doubled from 2818 to 5636.
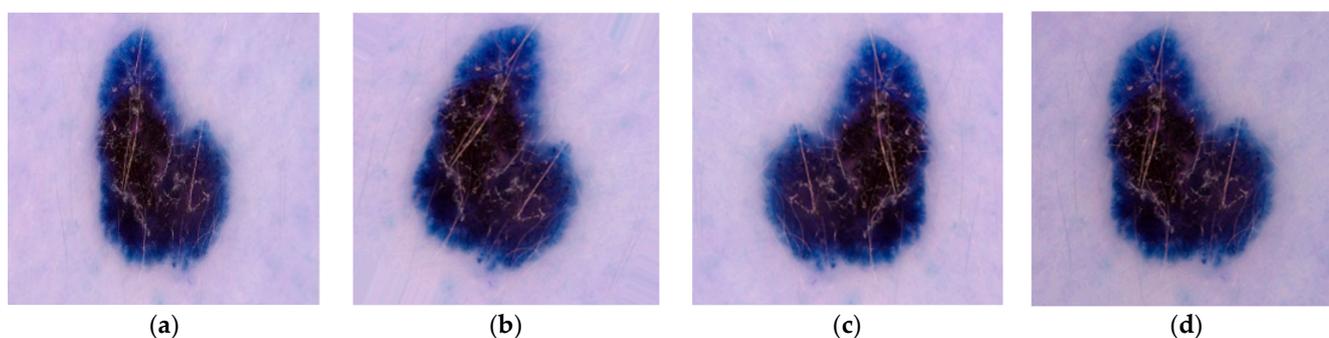

Figure 3. Data Augmentation Techniques in Sequence (**a**) Input Image, (**b**) Rotated Image, (**c**) Flipped Image, (**d**) Brightened Image.

Table 2 shows the total images of training, testing and validation data after augmentation. The augmentation is applied only on the training images. Previously, the training images of benign and malignant are 1534 and 1284, respectively. After the augmentation, there was a total of 5636 training images.

Table 2. Dataset Description of training images.

	Before Augmentation	After Augmentation
Benign	1534	3068
Malignant	1284	2568
Total	2818	5636

2.3. Feature Extraction Using VGG16

We found that the majority of biomedical imaging transfer learning methods used VGG approaches to achieve the highest levels of prediction accuracy after thoroughly analyzing the methodologies. The authors of this work are inspired to implement VGG16 [32] by hyper-tuning the parameters in order to achieve the maximum possible accuracy. This is a deep Convolutional Neural Network (CNN) architecture with several16 layers, known as VGG16. In ImageNet, the VGG16 model performs about 92.7 percent of the top-five tests correctly. There are more than 14 million photos in ImageNet [36], which can be divided into more than 1000 categories. It was also one of the most popular models submitted to the 2014 International Laser Sintering and Research Conference. In the VGG16 architecture, an input image with 224 * 224 size is applied as shown in Figure 4. The VGG16 architecture consists of five blocks. In the first and second block, two convolution layers (3 * 3), and one max pooling layer (2 * 2) are applied with 64 and 128 filters, respectively. In the third, fourth, and fifth blocks, three convolution layers with 256, 512, and 512 filters are used, respectively, followed by a max pool layer (2 * 2). Therefore, in the proposed work, the VGG16 model is further modified by adding one flatten layer, two dense layers with LeakyReLU activation function, and another dense layer with activation function (sigmoid) to enhance the accuracy of this model.
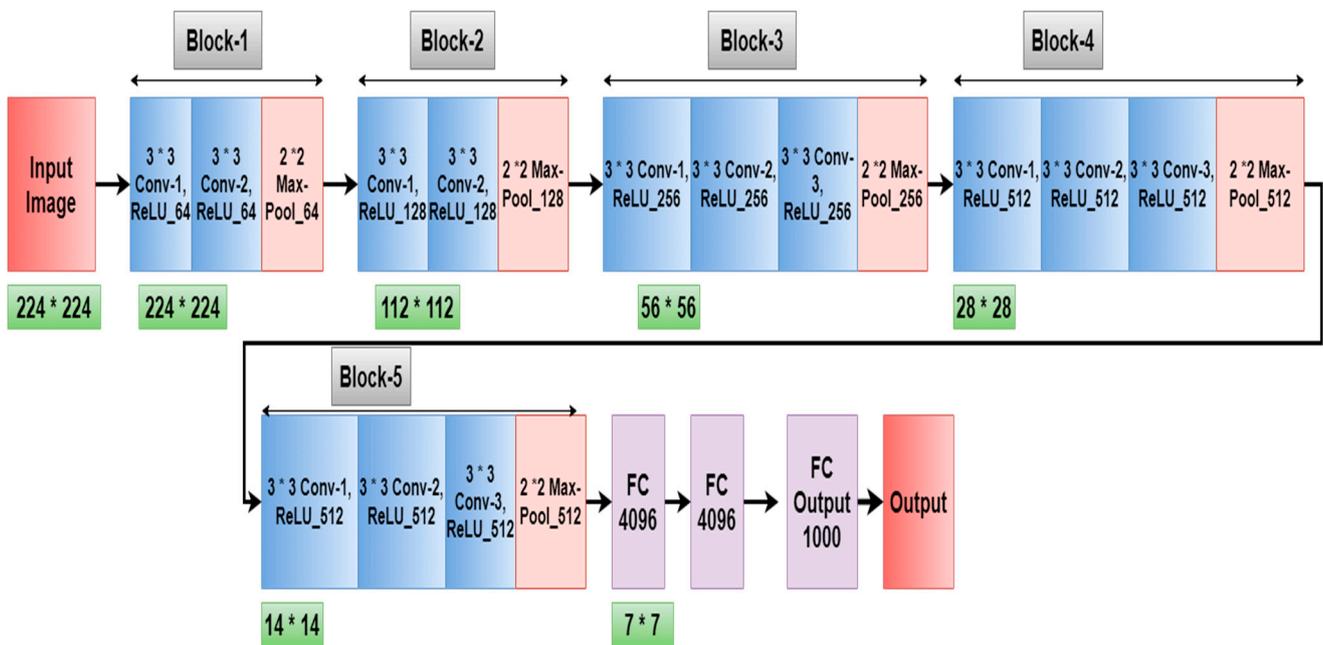

Figure 4. VGG16 Architecture.

Table 3 shows the images after filtration with each block after every conv layer and max-pool layer. Block 1 and block 2 consist of two conv layers and one max-pool layer. Hence the images of the 3rd convolution layer are not shown in block 1 and block 2. For each layer only a single filtered image is shown in the table. For example, in convolution layer 1 of block 1, 64 filters are used, so 64 filtered images will be received after that convolution layer.

Table 3. Filtered Images after Convolution and Maxpool Layer.

	Conv-1	Conv-2	Conv-3	Max Pool
Block 1			NA	
Block 2			NA	
Block 3				
Block 4				
Block 5				
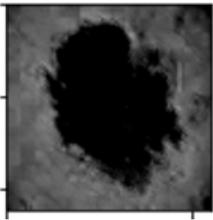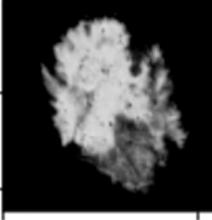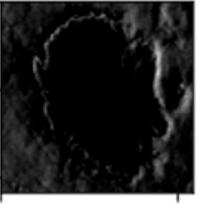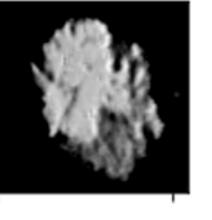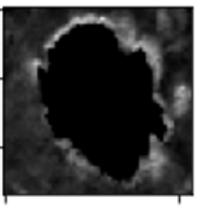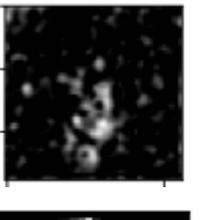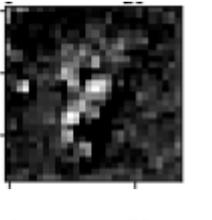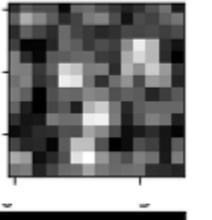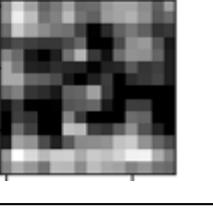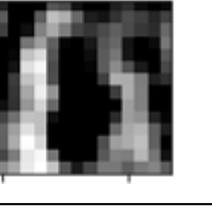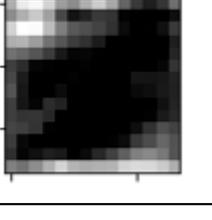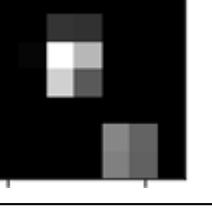
2.4. Fine Tuning of VGG16 Model

Figure 5 displays the fine tuning of the VGG16 model. Extracted features from the VGG16 model are provided as input to the flatten layer. After that, it is transferred to the two dense layers having 32 and 16 neurons, respectively, with LeakyReLU as the activation function. The third dense layer consists of two neurons and a sigmoid activation function. After that, the image is classified into one of two different classes of cancer (i.e., benign and malignant).
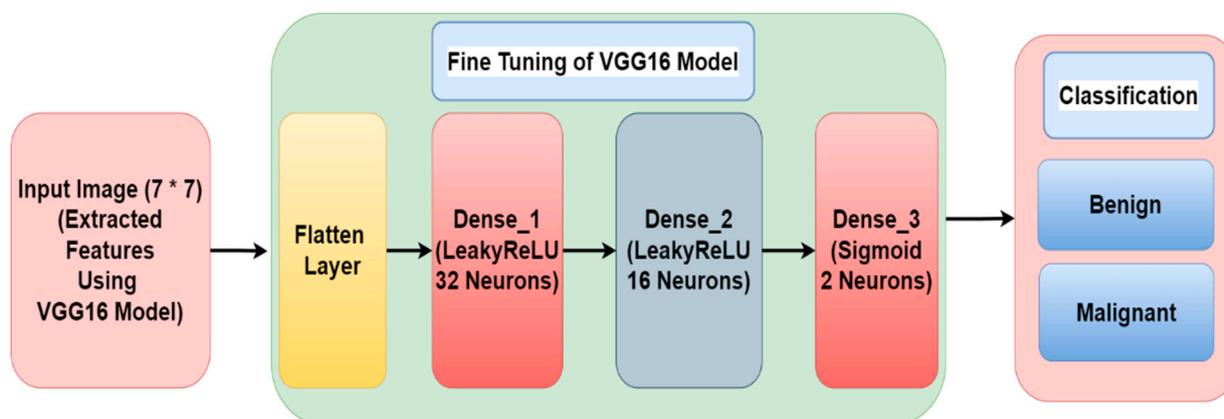
Figure 5. Fine Tuning of VGG16 Model.

Table 4 shows the parameters of the proposed model. The output image size after the VGG16 model is 7 * 7 and the number of parameters is 1,471,468. After VGG16, the flatten layer is added whose output shape is 25,088 * 1. After that, different layers like Dense_1, Dense_2 and Dense_3 are added having 802, 848, 528, and 17 layers, respectively. LeakyReLU and Sigmoid activation functions are used for these dense layers.

**Table 4.** Parameters of Proposed Model.

Layer	Shape of Output	Parameters
VGG16	7, 7, 512	1,471,468
Flatten	25,088	0
Dense_1	32	802,848
LeakyReLU_1	32	0
Dense_2	16	528
LeakyReLU_2	16	0
Dense_3	1	17
Total Parameters		15,518,081
Trainable Parameters		803,393
Non-trainable Parameters		14,714,688

3. Results and Discussion

This contains all of the outcomes obtained using the proposed model. On the Kaggle dataset, the model is tested. Various performance criteria such as precision, sensitivity, F1 Score, and accuracy are taken into account when analyzing the suggested model. An exploratory investigation is carried out using various hyper parameters, which are described in detail below.

3.1. Hyper Parameters Tuning

Different parameters such as optimizer [37], batch size and epochs are used for hyper parameters tuning on dermoscopy images. The Adam optimizer is a frequently used optimizer that has replaced the Stochastic Gradient Descent optimizer in terms of training the deep learning algorithms. Adam combines the various characteristics of RMSProp and AdaGrad optimizers. The expression of Adam optimizer is given in Equations (1) and (2):

$$p_t = \alpha_1 p_{t-1} + (1 - \alpha_1) \left[\frac{\delta L}{\delta w_t} \right] \tag{1}$$

$$q_t = \alpha_2 q_{t-1} + (1 - \alpha_2) \left[\frac{\delta L}{\delta w_t} \right]^2 \tag{2}$$

α_1 and α_2 are the decay rates, δL is loss function derivative, δw_t is weights derivative at t, wt signifies the weights, p_t is gradients collection, and q_t is past gradients sum of squares. Batch size is the most significant hyper parameter that is used for tuning any deep learning system. A large batch size causes the computational speedups during training of a deep learning model because of parallelism of GPUs but it may cause poor generalization. A small batch size causes faster convergence to good solutions. Hence there is always a competition between large and small batch size. In this paper, the proposed model is simulated with different batch sizes like 8, 16, 32, 64, and 128 to analyze which batch size will be suitable for better accuracy.

Epoch is the overall amount of times the whole dataset is received by neural network. When a model is trained for one epoch, it means that training dataset had one chance to update the internal parameters of the model. Therefore, the number of epochs should be more so that error can be minimized during learning of the model. But more epochs increase the computational time period. Hence, there should be a trade-off between a high and a small number of epochs. In this paper, the presented model is simulated using 10 and 20 epochs. Table 5 shows the name of hyper tuning parameters and their values.

Table 5. Hyper Tuning Parameters.

S. No.	Parameter	Value
1.	Batch size	8, 16, 32, 64, 128
2.	Optimizer	Adam
3.	Epochs	20

3.2. Model Accuracy and Model Loss Analysis

Training accuracy, validation accuracy for VGG16 and modified VGG16 is performed on the basis of model accuracy and model loss Figure 6 displays the graphs of training accuracy, validation accuracy for VGG16, and modified VGG16.

Figure 6a–e displays the accuracy for VGG16 and Figure 6f–j displays the training, validation accuracy of modified VGG16. The model is evaluated on 20 epochs. Figure 6a,f shows the graph of training and validation accuracy on 8 batch size for VGG16 and the modified VGG16 model, respectively.

It is observed that, for modified VGG16, the values of validation accuracy are more in comparison to the VGG16 model. The highest value is on the 11th epoch that is approximately 87% for the modified VGG16 model.

Figure 6b,g shows the graph of validation and training accuracy on 16 batch size for VGG16 and the modified VGG16 model, respectively. It is observed that, for modified VGG16, the values of validation accuracy are more in comparison to VGG16 model. The highest value is on the 6th epoch that is approximately 87% for the modified VGG16 model.

Figure 6c,h shows the graph of validation and training accuracy on 32 batch size for VGG16 and the modified VGG16 model, respectively. It is observed that, for modified VGG16, the values of validation accuracy are more in comparison to the VGG16 model. The highest value is on the 3rd and 18th epoch that is approximately 86% for the modified VGG16 model.

Figure 6d,i shows the graph of validation and training accuracy on 64 batch size for VGG16 and the modified VGG16 model, respectively. It is observed that, for modified VGG16, the values of validation accuracy are more in comparison to the VGG16 model. The highest value is on the 11th epoch that is 86% for the modified VGG16 model.

Figure 6e,j shows the graph of validation and training accuracy on 128 batch size for VGG16 and the modified VGG16 model, respectively. It is observed that, for modified VGG16, the values of validation accuracy are more in comparison to the VGG16 model. The highest value is on the 18th epoch that is approximately 87.5% for the modified VGG16 model. It can be analyzed from Figure 6 that, for all the batch sizes as well as each epoch, validation accuracy is better for modified VGG16 as compared to VGG16. In any deep

learning model, training and validation loss reduces as the number of epochs increases. Starting from 0 to 20 epoch values, peak points are shown in all the figures.

Figure 7 shows the graphs of loss for VGG16 and modified VGG16. Figure 7a–e shows the training loss for VGG16, and Figure 7f–j shows the training loss, validation loss for modified VGG16. The values of training loss are compared to validation loss. The model is evaluated on 20 epochs. Figure 7a,f shows the graph of training and validation loss on 8 batch size for VGG16 and the modified VGG16 model, respectively. It is observed that, for modified VGG16, the values of validation loss are less in comparison to the VGG16 model. Similarly, for all the batch sizes modified VGG16 is showing better results as compared to VGG16 in terms of validation loss. Starting from 0 to 20 epoch values, peak points are shown in all the figures.

Batch Size	VGG16	Modified VGG16
8	(a)	(f)
16	(b)	(g)
32	(c)	(h)

Figure 6. *Cont.*

64

(**d**)


(**i**)

128

(**e**)

(**j**)
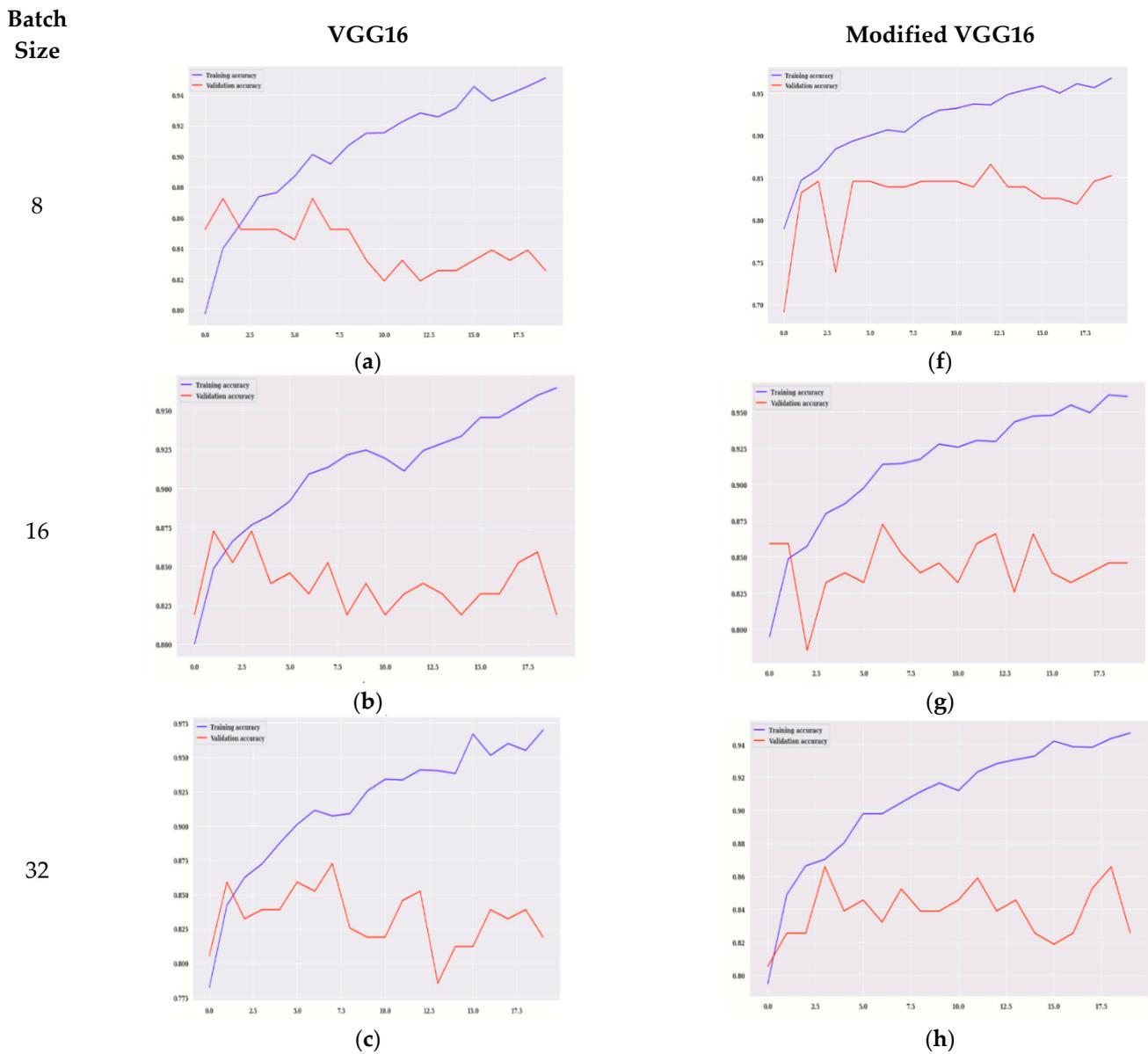
Figure 6. Training Accuracy for VGG16 on Batch size (**a**) 8 (**b**) 16 (**c**) 32 (**d**) 64 (**e**) 128, Training Accuracy for Modified VGG16 on Batch size (**f**) 8 (**g**) 16 (**h**) 32 (**i**) 64 (**j**) 128.

Batch Size	**VGG16**	**Modified VGG16**

8

(**a**)

(**f**)

16

(**b**)

(**g**)
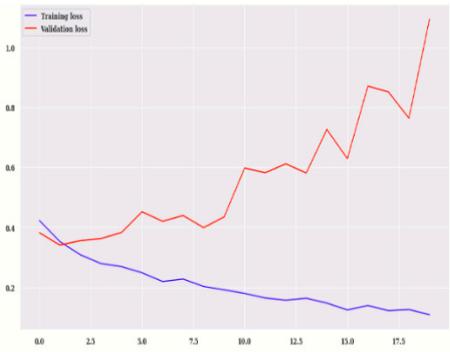
Figure 7. *Cont.*
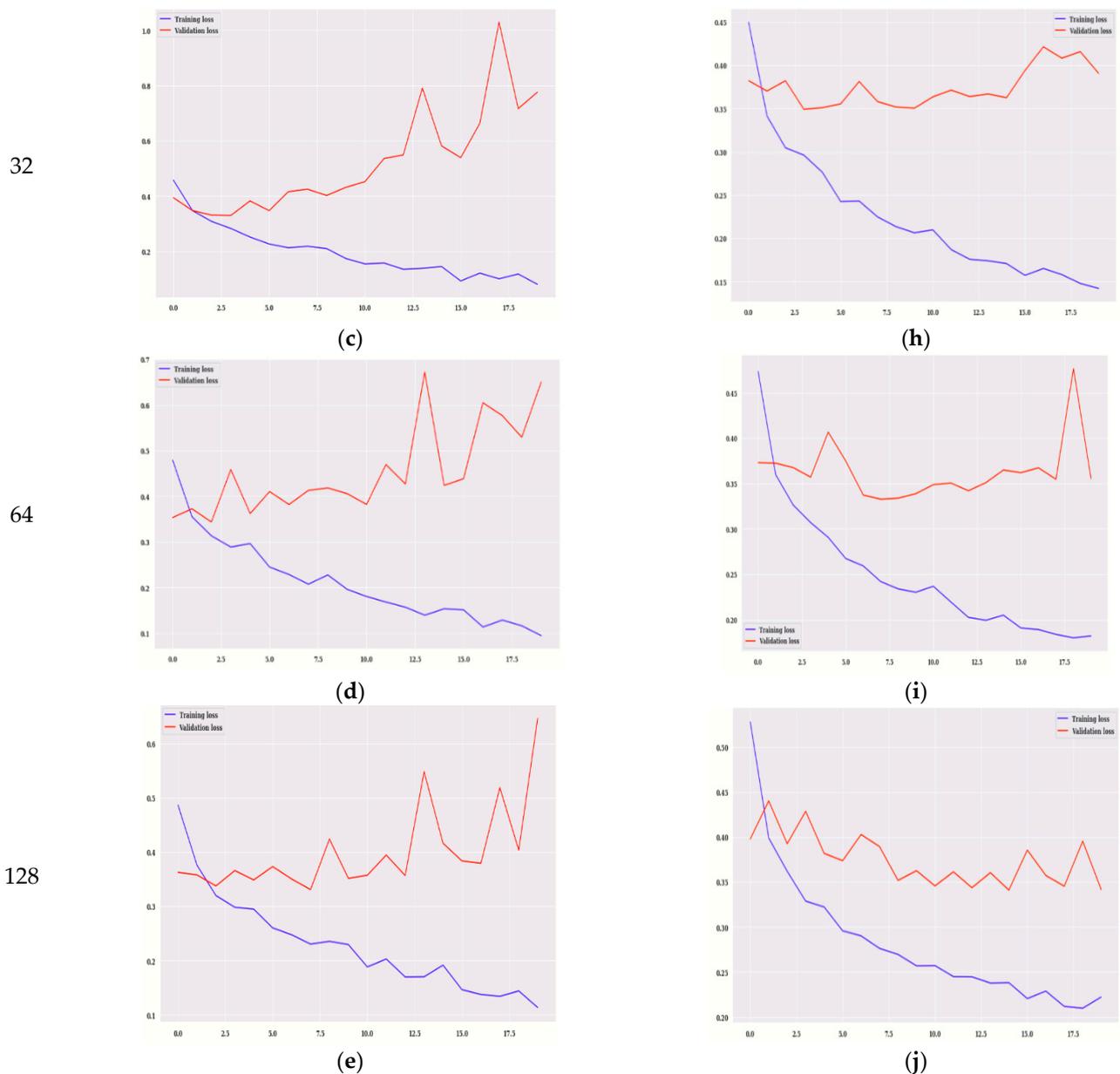
Figure 7. Training Loss for VGG16 on Batch Size (**a**) 8 (**b**) 16 (**c**) 32 (**d**) 64 (**e**) 128, Training Loss for Modified VGG16 (**f**) 8 (**g**) 16 (**h**) 32 (**i**) 64 (**j**) 128.

Table 6 shows the loss and accuracy for training and validation of the modified VGG16 model for two different epoch values, 10 and 20. The model is simulated with the Adam optimizer at five different batch sizes (i.e., 8, 16, 32, 64, and 128). It can be seen that, on the 10th epoch, the training accuracy is maximum at batch size 32, which is 0.9255 whereas training loss is minimum (i.e., 0.1736). Whereas, on the same epoch value, validation accuracy is maximum (i.e., 84.56%), and the validation loss is less (i.e., 0.3514) for batch size 32. From Table 6, it is also analyzed that, on 20th epoch, the value of training accuracy is maximum at batch size 32, which is 0.9698, whereas training loss is minimum (i.e., 0.0801). Whereas, on the same epoch, value validation accuracy is maximum (i.e., 82.55%) at batch size 8, and the validation loss is minimum (i.e., 0.6468) at batch size 128.

Overall, it can be concluded from this table that, on the 10th and 20th epoch, the training results are best at batch size 32, whereas, validation results are best on the 10th epoch at batch size 128.

Table 6. Training Performance of Modified Vgg16 Model with Adam Optimizer.

Epoch Value	Batch Size	Train Loss	Train Accuracy	Validation Loss	Val Accuracy (%)
	8	0.1912	0.9150	0.4346	83.22
	16	0.1705	0.9246	0.4664	83.89
10	32	0.1736	0.9255	0.4310	81.88
	64	0.1957	0.9168	0.4034	83.22
	128	0.2293	0.8971	0.3514	84.56
	8	0.1077	0.9510	1.0940	82.55
	16	0.0876	0.9645	0.7804	81.88
20	32	0.0801	0.9698	0.7754	81.88
	64	0.0944	0.9634	0.6499	80.54
	128	0.1133	0.9546	0.6468	79.87

3.3. Confusion Matrix

The confusion matrix provided predictions of true and false values as shown in Figure 8. True labels are indicated vertically, and predicted labels are indicated horizontally from which False Negatives (FN), False Positives (FP), True Positives (TP), and True Negatives (TN) can be calculated. The parameter accuracy is calculated using TP, TN, FP, and FN as given in Equation (3)

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \tag{3}$$

Figure 8a–e shows the confusion matrix for VGG16, and Figure 8f–j shows the confusion matrix for modified VGG16. In the case of modified VGG16, the values of accuracy are higher in comparison to the VGG16 model. On batch size 8, the accuracy on the modified VGG16 model is 86.67%, whereas, on the VGG16 model, it is 85.45%. On batch size 16, the accuracy is approximately similar to the VGG16 model and the modified VGG16 model as shown in Figure 8b,g. with batch size 32, the accuracy on the modified VGG16 model is better (i.e., 88.18%), whereas, on the VGG16 model, it is 84.24%. On batch size 64, the accuracy on the modified VGG16 model is 87.58%, whereas, on the VGG16 model, it is 86.67%. On batch size 128, the accuracy on the modified VGG16 model is 89.09%, whereas, on the VGG16 model, it is 82.42%. From Figure 8, it can be concluded that the modified VGG16 model is showing better accuracy as compared to VGG16 for each batch size (i.e., 8, 16, 32, 64, and 128).

3.4. Confusion Matrix Parameter Analysis

The confusion matrix parameter values are calculated by using Equations (4), (5), and (6), respectively.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5}$$

$$\text{F1 Score} = \frac{2 \times \text{TP}}{2 \times (\text{TP} + \text{FP} + \text{FN})} \tag{6}$$

Figure 9 displays the values of confusion matrix parameter analysis for the benign and malignant class. Figure 9a displays precision for benign and malignant disease class on the VGG16 model and the modified VGG16 model. From the figure, it can be seen that the values of precision are higher in case of the modified VGG16 model at each batch size for benign as well as malignant class. For benign class, modified VGG16 is working best for batch size 8, whereas, for malignant class, it is working best for batch size 8 and 128.

**Batch Size**

VGG16

Modified VGG16

(**a**)

(**f**)

(**b**)

(**g**)

(**c**)

(**h**)

Figure 8. *Cont.*

(**d**)

(**i**)


(**e**)

(**j**)
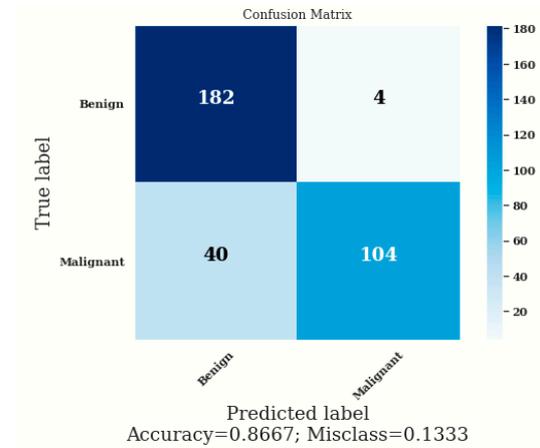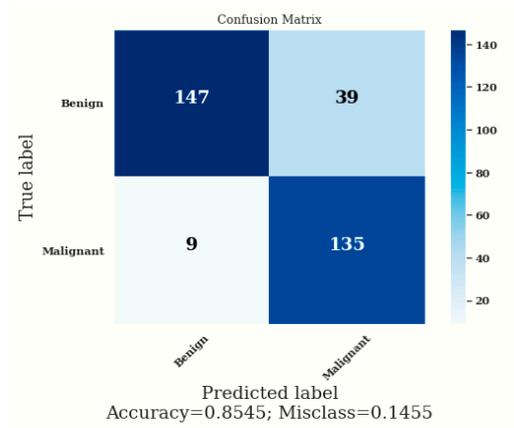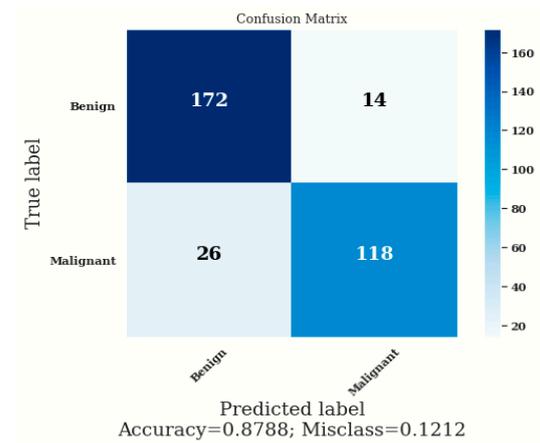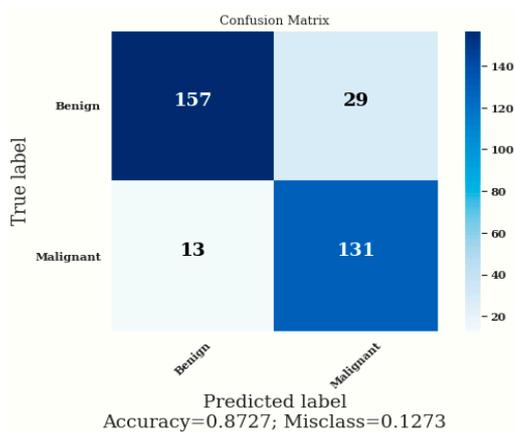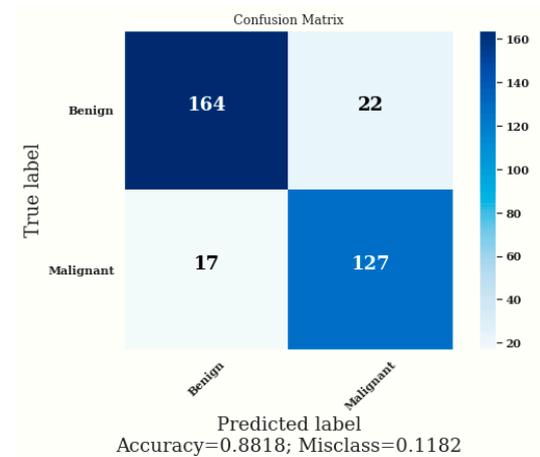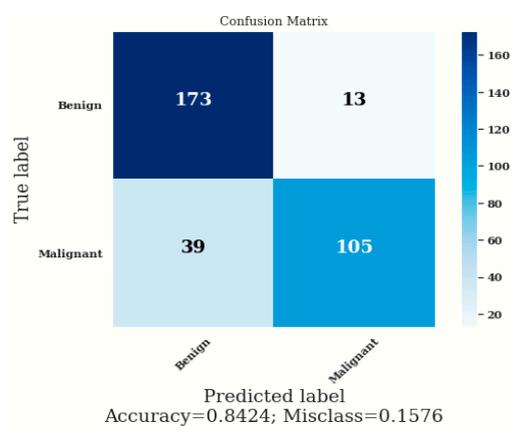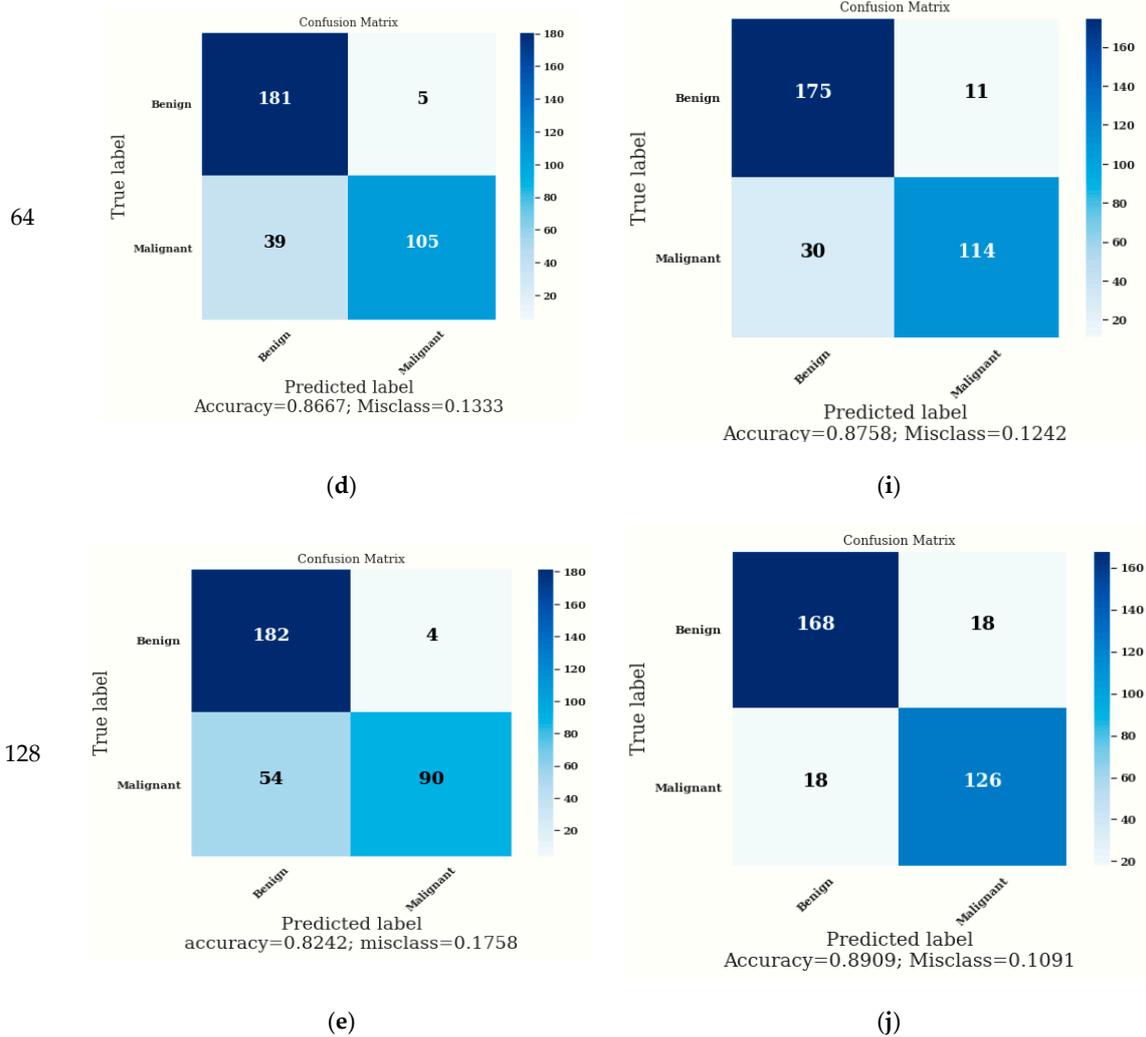
Figure 8. Confusion Matrix for VGG16 on Batch Size (**a**) 8 (**b**) 16 (**c**) 32 (**d**) 64 (**e**) 128, Confusion Matrix for Modified VGG16 on Batch Size (**f**) 8 (**g**) 16 (**h**) 32 (**i**) 64 (**j**) 128.

Figure 9b shows the values of sensitivity for the benign and malignant disease classes on the VGG16 model and the modified VGG16 model. From the figure, it can be seen that the values of sensitivity are higher in the case of the modified VGG16 model at each batch size for benign as well as malignant class. For benign class, modified VGG16 is working best for batch size 8 and 128 whereas, for malignant class, it is working best for batch size 8 and 16.

Figure 9c shows the values of F1 score for VGG16 and the modified VGG16 model. The values of F1 Score are higher in the case of the modified VGG16 model at each batch size for benign as well as malignant class as compared to the VGG16 model. For benign class, modified VGG16 is working best for batch size 16, 64 and 128 whereas; for malignant class, it is working best for batch size 128.

Figure 9d shows the accuracy values of benign and malignant disease class on VGG16 and the modified VGG16 model. From the figure, it is detected that the overall ac-curacy is high in the case of the modified VGG16 model at each batch size as compared to the VGG16 model. Modified VGG16 is working best for batch size 128 in terms of overall accuracy, and the value is 89.09%.
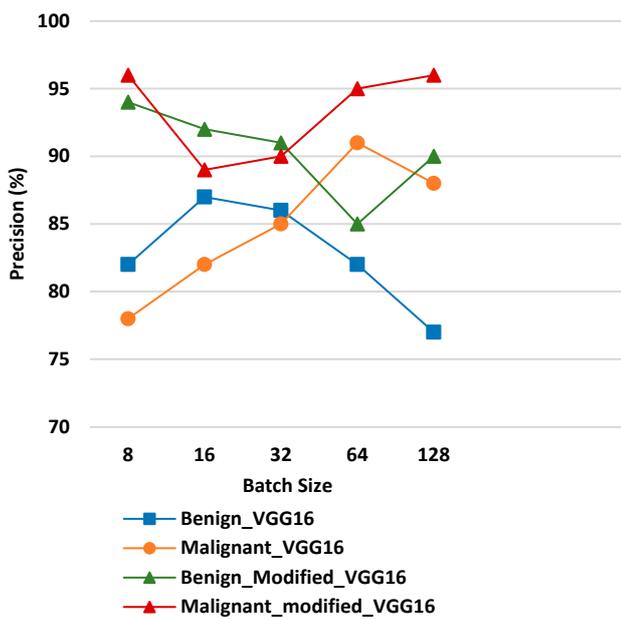
(**a**) Precision
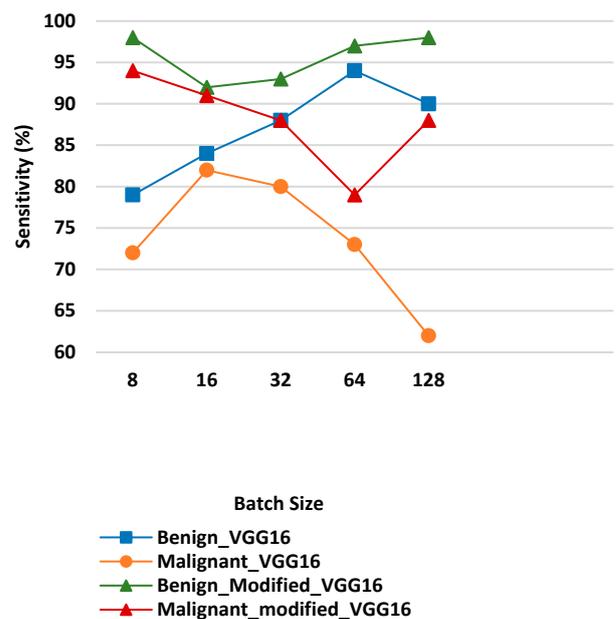
(**b**) Sensitivity
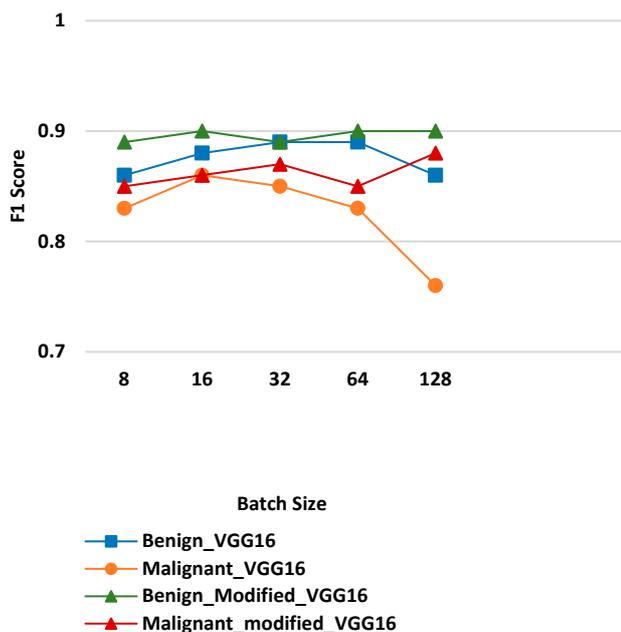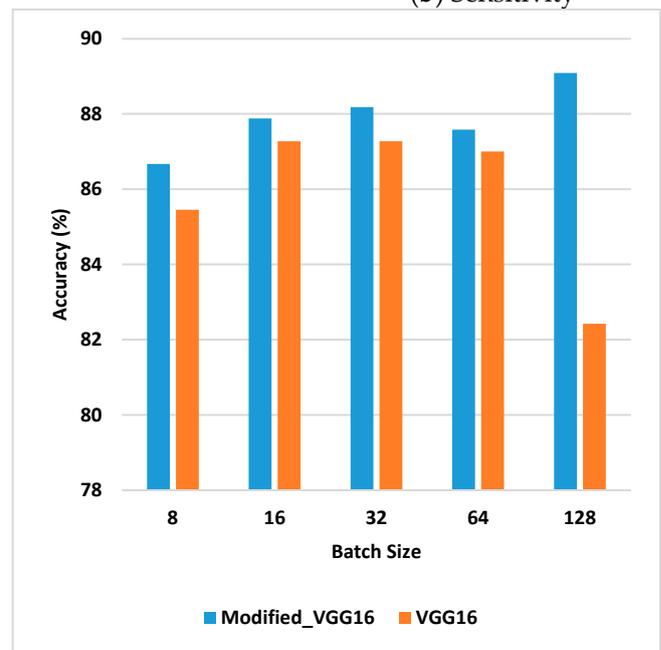
(**c**) F1 Score

(**d**) Accuracy

Figure 9. Confusion Matrix Parameters (**a**) Precision, (**b**) Sensitivity, (**c**) F1 Score, (**d**) Accuracy.

3.5. Comparison with the State-of-the-Art

A comparison with other state-of-the-art methods has been performed using skin dermoscopy images in terms of accuracy and is presented in Table 7. The result analysis shows that the presented model has achieved good accuracy as compared to other state-of-the-art models. The accuracy is different in all studies, as different datasets (HAM10000, Kaggle and clinical images) are used. With HAM10000 dataset, Khan et al. [8] achieved an accuracy of 80.46% using VGG16 model architecture, and Agrahari et al. [10] and Chaturvedi et al. [11] had achieved accuracy rates of 80.81% and 83.10%, respectively on MobileNet model architecture.

Table 7. Comparison of the Proposed Model with State-of-the-Art Techniques.

Ref	Technique Used	Dataset	Accuracy (%)
Khan et al. [8]	VGG16	HAM10000	80.46
Agrahari et al. [10]	MobileNet	HAM10000	80.81
Chaturvedi et al. [11]	MobileNet	HAM10000	83.10
Abdar et al. [13]	Bayesian Deep Learning Method	Kaggle	88.95
Fjisawa et al. [14]	Deep Convolutional Neural Network	Clinical Images	76.50
Garcia et al. [15]	Machine Learning	Interactive Atlas of Dermoscopy	88.00
Hasan et al. [38]	CNN	Kaggle	89.5
Singh et al. [39]	ResNet50	Kaggle	80.3
Proposed	**Modified VGG16 architecture**	**Kaggle**	**89.09**

4. Conclusions

The modified VGG16 model is trained using transfer learning. The training and testing of the model is performed on the Kaggle dataset. The presented model has been analyzed with various batch sizes of 8, 16, 32, 64, and 128 using the Adam optimizer and 10 Epochs. The proposed model is working best with overall accuracy of 89.09% on 128 batch size with Adam optimizer and 10 epochs. There is still a scope in improving the overall accuracy of the presented model. It can be enhanced by increasing both true positives as well as true negatives simultaneously. There is always a possibility to build a more suitable model for detection of skin cancer.

Author Contributions: Formal analysis, V.A.; Funding acquisition, A.A.; Methodology, S.G.; Software, S.R.N.; Writing—original draft, R.C.P.; Writing—review & editing, A.K.J.S. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bauer, A.; Adam, K.E.; Soyer, P.H.; Adam, K.W.J. Prevention of Occupational Skin Cancer. In *Kanerva's Occupational Dermatology*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 1685–1697.
2. Al-antari, M.A.; Rivera, P.; Al-masni, M.; Añazco, V.; Gi, E.; Kim, G.; Park, T.-Y.M.; Kim, T.-S.H. An Automatic Recognition of Multi-class Skin Lesions via Deep Learning Convolutional Neural Networks. In Proceedings of the ISIC2018: Skin Image Analysis Workshop and Challenge, Granada, Spain, 20 September 2018.
3. Seeley, R.; Stephens, D.; Philip, T. *Anatomy and Physiology*; McGraw-Hill: New York, NY, USA, 2008; pp. 1–1266.
4. Naylor, P.; Laé, M.; Reyal, F.; Walter, T. Nuclei segmentation in histopathology images using deep neural networks. In Proceedings of the 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), Melbourne, VIC, Australia, 18–21 April 2017; pp. 933–936.
5. Haenssle, H.A.; Fink, C.; Schneiderbauer, R.; Toberer, F.; Buhl, T.; Blum, A.; Kalloo, A.; Hassen, A.B.H.; Thomas, L.; Enk, A.; et al. Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann. Oncol.* **2018**, *29*, 1836–1842. [CrossRef] [PubMed]
6. Dorj, U.O.; Lee, K.K.; Choi, J.Y.; Lee, M. The skin cancer classification using deep convolutional neural network. *Multimed. Tools Appl.* **2018**, *77*, 9909–9924. [CrossRef]
7. Han, S.S.; Kim, M.S.; Lim, W.; Park, G.H.; Park, I.; Chang, S.E. Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. *J. Investig. Dermatol.* **2018**, *138*, 1529–1538. [CrossRef]

8. Khan, M.D.; Uddin, A.H.; Nahid, A.A.; Bairagi, A.K. Skin Cancer Detection from Low-Resolution Images Using Transfer Learning. In *Intelligent Sustainable Systems*; Springer: Singapore, 2022; pp. 317–334.

9. Mohakud, R.; Dash, R. Skin cancer image segmentation utilizing a novel EN-GWO based hyper-parameter optimized FCEDN. *J. King Saud Univ.-Comput. Inf. Sci.* 2022, *in press*. [CrossRef]

10. Agrahari, P.; Agrawal, A.; Subhashini, N. Skin Cancer Detection Using Deep Learning. In *Futuristic Communication and Network Technologies*; Springer: Singapore, 2022; pp. 179–190.

11. Chaturvedi, S.S.; Gupta, K.; Prasad, P.S. Skin lesion analyser: An efficient seven-way multi-class skin cancer classification using mobilenet. In Proceedings of the International Conference on Advanced Machine Learning Technologies and Applications, Cairo, Egypt, 20–22 March 2020; Springer: Singapore, 2020; pp. 165–176.

12. Hosny, K.M.; Kassem, M.A.; Foaud, M.M. Skin cancer classification using deep learning and transfer learning. In Proceedings of the 2018 9th Cairo International Biomedical Engineering Conference (CIBEC), Cairo, Egypt, 20–22 December 2018; pp. 90–93.

13. Abdar, M.; Samami, M.; Mahmoodabad, S.D.; Doan, T.; Mazoure, B.; Hashemifesharaki, R.; Liu, L.; Khosravi, A.; Acharya, U.R.; Makarenkov, V.; et al. Uncertainty quantification in skin cancer classification using three-way decision-based Bayesian deep learning. *Comput. Biol. Med.* **2021**, *135*, 104418. [CrossRef] [PubMed]

14. Fujisawa, Y.; Otomo, Y.; Ogata, Y.; Nakamura, Y.; Fujita, R.; Ishitsuka, Y.; Watanabe, R.; Okiyama, N.; Ohara, K.; Fujimoto, M. Deep-learning-based, computer-aided classifier developed with a small dataset of clinical images surpasses board-certified dermatologists in skin tumour diagnosis. *Br. J. Dermatol.* **2019**, *180*, 373–381. [CrossRef]

15. Garcia-Arroyo, J.L.; Garcia-Zapirain, B. Recognition of pigment network pattern in dermoscopy images based on fuzzy classification of pixels. *Comput. Methods Programs Biomed.* **2018**, *153*, 61–69. [CrossRef]

16. Iyatomi, H.; Oka, H.; Celebi, M.E.; Ogawa, K.; Argenziano, G.; Soyer, H.P.; Tanaka, M. Computer-based classification of dermoscopy images of melanocytic lesions on acral volar skin. *J. Investig. Dermatol.* **2008**, *128*, 2049–2054. [CrossRef]

17. Chatterjee, S.; Dey, D.; Munshi, S. Optimal selection of features using wavelet fractal descriptors and automatic correlation bias reduction for classifying skin lesions. *Biomed. Signal Process. Control* **2018**, *40*, 252–262. [CrossRef]

18. Chatterjee, S.; Dey, D.; Munshi, S. Integration of morphological preprocessing and fractal based feature extraction with recursive feature elimination for skin lesion types classification. *Comput. Methods Programs Biomed.* **2019**, *178*, 201–218. [CrossRef]

19. González-Díaz, I. Dermaknet: Incorporating the knowledge of dermatologists to convolutional neural networks for skin lesion diagnosis. *IEEE J. Biomed. Health Inform.* **2018**, *23*, 547–559. [CrossRef] [PubMed]

20. Kawahara, J.; Daneshvar, S.; Argenziano, G.; Hamarneh, G. Seven-point checklist and skin lesion classification using multitask multimodal neural nets. *IEEE J. Biomed. Health Inform.* **2018**, *23*, 538–546. [CrossRef] [PubMed]

21. Koohbanani, N.A.; Jahanifar, M.; Tajeddin, N.Z.; Gooya, A.; Rajpoot, N. Leveraging transfer learning for segmenting lesions and their attributes in dermoscopy images. *arXiv* 2018, arXiv:1809.10243.

22. Filali, Y.; El Khoukhi, H.; Sabri, M.A.; Yahyaouy, A.; Aarab, A. Texture Classification of skin lesion using convolutional neural network. In Proceedings of the 2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS), Fez, Morocco, 3–4 April 2019; pp. 1–5.

23. Kadampur, M.A.; Al Riyaee, S. Skin cancer detection: Applying a deep learning based model driven architecture in the cloud for classifying dermal cell images. *Inform. Med. Unlocked* **2020**, *18*, 100282. [CrossRef]

24. Menegola, A.; Fornaciali, M.; Pires, R.; Bittencourt, F.V.; Avila, S.; Valle, E. Knowledge transfer for melanoma screening with deep learning. In Proceedings of the International Symposium on Biomedical Imaging, Melbourne, Australia, 18–21 April 2017.

25. Available online: https://www.kaggle.com/fanconic/skin-cancer-malignant-vs-benign (accessed on 8 February 2021).

26. Lee, H.D.; Mendes, A.I.; Spolaor, N.; Oliva, J.T.; Parmezan AR, S.; Wu, F.C.; Fonseca-Pinto, R. Dermoscopic assisted diagnosis in melanoma: Reviewing results, optimizing methodologies and quantifying empirical guidelines. *Knowl.-Based Syst.* **2018**, *158*, 9–24. [CrossRef]

27. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]

28. Duggani, K.; Nath, M.K. A Technical Review Report on Deep Learning Approach for Skin Cancer Detection and Segmentation. *Data Anal. Manag.* **2021**, *54*, 87–99.

29. Khan, M.A.; Zhang, Y.D.; Sharif, M.; Akram, T. Pixels to Classes: Intelligent Learning Framework for Multiclass Skin Lesion Localization and Classification. *Comput. Electr. Eng.* **2021**, *90*, 106956. [CrossRef]

30. Alharithi, F.; Almulihi, A.; Bourouis, S.; Alroobaea, R.; Bouguila, N. Discriminative Learning Approach Based on Flexible Mixture Model for Medical Data Categorization and Recognition. *Sensors* **2021**, *21*, 2450. [CrossRef]

31. Masud, M.; Singh, P.; Gaba, G.S.; Kaur, A.; Alghamdi, R.A.; Alrashoud, M.; Alqahtani, S.A. CROWD: Crow Search and Deep Learning based Feature Extractor for Classification of Parkinson's Disease. *ACM Trans. Internet Technol. (TOIT)* **2021**, *21*, 1–18. [CrossRef]

32. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.

33. Rodrigues, D.d.A.; Ivo, R.F.; Satapathy, S.C.; Wang, S.; Hemanth, J.; Reboucas Filho, P.P. A new approach for classification skin lesion based on transfer learning, deep learning, and IoT system. *Pattern Recognit. Lett.* **2020**, *136*, 8–15. [CrossRef]

34. Moataz, L.; Salama, G.I.; Abd Elazeem, M.H. Skin Cancer Diseases Classification using Deep Convolutional Neural Network with Transfer Learning Model. *J. Phys. Conf. Ser.* **2021**, *2128*, 012013. [CrossRef]

35. Ashim, L.K.; Suresh, N.; Prasannakumar, C.V. A Comparative Analysis of Various Transfer Learning Approaches Skin Cancer Detection. In Proceedings of the 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 3–5 June 2021; pp. 1379–1385.
36. Nayak, S.R.; Nayak, D.R.; Sinha, U.; Arora, V.; Pachori, R.B. Application of deep learning techniques for detection of COVID-19 cases using chest X-ray images: A comprehensive study. *Biomed. Signal Process. Control* **2020**, *64*, 102365. [CrossRef] [PubMed]
37. Landro, N.; Gallo, I.; La Grassa, R. Mixing ADAM and SGD: A Combined Optimization Method. *arXiv* **2020**, arXiv:2011.08042.
38. Hasan, M.; Barman, S.D.; Islam, S.; Reza, A.W. Skin cancer detection using convolutional neural network. In Proceedings of the 2019 5th International Conference on Computing and Artificial Intelligence, Bali, Indonesia, 17–20 April 2019; pp. 254–258.
39. Singh, V.; Nwogu, I. Analyzing skin lesions in dermoscopy images using convolutional neural networks. In Proceedings of the 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Miyazaki, Japan, 7–10 October 2018; pp. 4035–4040.