2022 25th International Conference on Computer and Information Technology (ICCIT)
17-19 December, Cox's Bazar, Bangladesh

A CNN Based Approach to Classify Skin Cancers using Transfer Learning

Ayasha Hossain Jui*, Samia Sharnami† and Aminul Islam‡
Computer Science and Engineering Discipline, Khulna University
Khulna, Bangladesh
Email: *jui1801@cseku.ac.bd, †sharna1833@cseku.ac.bd, ‡aminul@cse.ku.ac.bd

***Abstract*—Skin cancer is one of the most frequent types of cancer in the world. Malignant skin cancers can become life-threatening if not treated in an early stage. Some skin cancers like Squamous Cell Carcinoma (SCC), Melanoma (MEL), and Basal Cell Carcinoma (BCC) are always malignant and can cause fatal damage to the skin. Therefore, early identification is essential for minimizing the harm. This study uses transfer learning and explores the famous pre-trained model, Xception, with necessary fine tuning to classify eight types of skin cancers. During the training of the proposed model, data augmentation techniques are used to introduce diversity in the training phase. The proposed method is tested on the International Skin Imaging Collaboration (ISIC) 2019 dataset. It exhibits better results than similar studies which can help early detection of malignant skin cancers.**

***Index Terms*—Skin Cancer, Data Augmentation, Feature extraction, Fine-tuning, Transfer Learning, Convolutional Neural Network**

I. Introduction

Skin is the largest organ of the human body. It has three layers: epidermis, dermis, and hypodermis. Among these three skin layers, the epidermis is exposed to the outer environment, which causes different skin diseases, and skin cancer is one of them. The epidermis contains melanocytes, which control the pigment in our skin. Skin cancers occur when the melanocyte cells grow out of control. The main cause of this disease is too much exposure to ultraviolet (UV) rays which emit from the sun, tanning beds, sunlamps etc. Skin cancer can affect anyone regardless of age, skin tone, or gender. However, an estimation says that Caucasian men over 50 have the highest risk of getting skin cancer. Also, mostly the people with less pigment (melanin) in their skin are at higher risk [1]. There are many types of skin cancers, such as Squamous Cell Carcinoma (SCC), Melanoma (MEL), Dermatofibroma (DF), Melanocytic Nevus (NV), Basal Cell Carcinoma (BCC), Actinic Keratosis (AK), Vascular lesion (VASC), Benign Keratosis (BKL), etcetera. The remaining skin malignancies are benign, except for vascular lesions which can be either malignant or benign. Basal cell carcinoma, melanoma, and squamous cell carcinoma are always malignant [2]. Squamous cell carcinoma can develop from actinic keratosis over time. Melanoma is third in terms of frequency among skin cancers worldwide, after basal cell carcinoma and squamous cell carcinoma. Skin cancer can develop anywhere on the body, but it is most common on skin exposed to ultraviolet radiation from the sun, tanning beds or lamps.

The lifetime risk of developing melanoma for white people is approximately 2.6% (1 in 38), 0.1% (1 in 1,000) for Black people, and 0.6% (1 in 167) for Hispanic people [3]. In 2020, there were an estimated 324.6 thousand new cases of melanoma skin cancer and almost 57 thousand deaths due to melanoma skin cancer worldwide [3]. In the US, melanoma is the fifth most prevalent cancer in males and the seventh most common disease in women. Unfortunately, the incidence of melanoma skin cancer is increasing every year [4]. The incidence rate and mortality rate of melanoma have been increasing worldwide for the last few decades [5].

Skin cancer can be diagnosed by physical examination or biopsy (by removing a tissue sample from the affected area for testing). Physical examination may give a false diagnosis. On the other hand, biopsy is costly and time-consuming. So, a method is proposed to reduce this problem by introducing a computer-aided solution by visual inspection. The risk of damage can be decreased by detecting skin cancer early, but it can cause fatal damage if found too late.

Skin cancer detection takes about two to three weeks to get the results from a biopsy which is a very long time. It can take a deadly form within this time period. Because, it can become life-threatening in just six weeks. If it is not treated in the meantime, it can spread to other body parts. For this reason, a computer-aided system is needed to reduce the detection time. Several researchers have already proposed their models for skin cancer classification. As datasets are few, it is difficult to identify the best model. So, a large dataset and appropriate learning technique is necessary to classify skin cancers. The proposed method has improved classification accuracy by applying transfer learning on ISIC 2019 Skin Cancer Dataset.

II. Related Works

Some related works done earlier by the researchers are discussed in this section. Many researchers have proposed Convolutional Neural Network (CNN) base models for classifying skin cancers. The authors of [6] have implemented four CNN based models, which are SENet154, InceptionResNetV2, InceptionV4, and PNASNet-5-Large, and an ensemble of all those. The ISIC-2018 dataset is used to test the models, and

979-8-3503-4602-2/22/$31.00 ©2022 IEEE

before the image data is fed into the network, it is pre-processed. They have augmented the dataset to solve the imbalance dataset problem. In a dataset, when the classes have significant distribution differences, then the dataset is called an imbalanced dataset. In this paper, authors have achieved an accuracy of 73% after ensembling the four models, while PNASNet-5-Large has the highest result of 76% among the five techniques. On the other hand, in [7], the authors have fine-tuned three different pre-trained models, such as MobileNet-50, DenseNet-121 and ResNet50, and an ensemble of all the models and tested them on the ISIC-2018 dataset. They have back propagated the weighted loss from the loss layer and augmented the image data to tackle the data imbalance problem. By solving the imbalanced dataset problem without image pre-processing, they achieved higher accuracy (77.5%) by ensembling the three models.

For skin lesion classification, two pre-trained models, MobileNet and VGG-16, have been implemented by the authors of [8]. They proposed a custom-made Convolutional Neural Network (CNN) model with five layers, and all were tested on the HAM10000 dataset. The paper shows that the custom CNN model has achieved an accuracy of 80.61% from 83.152% after augmenting the dataset. The accuracy of one of the single models has also decreased after data augmentation. In [9], the authors have used skin segmentation algorithms to improve the test results. They have proposed a custom CNN model with nine layers and achieved 80.52% accuracy without data augmentation. Without augmenting the dataset, the authors of [10] have improved the test results to 82.56% by using a different lesion segmentation technique. In this paper, the entropy-based weighting and the first-order cumulative moment (EW-FCM) are calculated, and the results are used to segment the lesion area from its background. Then, seven different skin lesions are classified, and the deep learning model wide-ShuffleNet is used for feature extraction.

In [11], the authors have implemented two fine-tuned, pre-trained models, such as VGG16 and GoogLeNet, and an ensemble of the two models. They have used max-RGB for colour normalization and transfer learning on the models to improve the results. The results have improved after combining the models. They have achieved 81.5% accuracy after applying the ensemble model. On the contrary, the authors of [12] have achieved 80.1% accuracy after ensembling six pre-trained models, such as VGG19, ResNet50, InceptionV3, DenseNet121, Xception, and VGG16. They have used data augmentation and a different technique for augmenting the data in which weights are applied to the minority or important classes, such as 'malignant'.

The authors of [13] have proposed to design the WonDerM pipeline, which consists of the following steps, such as re-sampling the pre-processed skin lesion images, building neural network architectures fine-tuned with segmentation task data, and using an ensemble method to classify seven types of skin diseases. For classification, the DenseNet model architecture is used. Hair removal and data augmentation techniques are implemented to improve the output results. The authors have achieved 78.5% test accuracy. Rathod et al. [14] have proposed a CNN model to extract features from dermoscopic images. The authors have achieved 70% training accuracy without image augmentation and fine-tuning the models. They have used the softmax classifier for multiple skin lesion classification. Image pre-processing techniques such as noise removal and image enhancement are used to improve output results.

III. Proposed Classification System

A skin cancer classification system is proposed in this study which includes image augmentation, feature extraction, transfer learning and model fine-tuning.

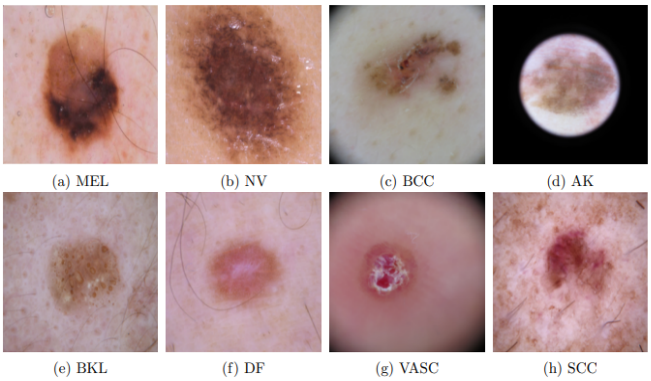

Fig. 1: Eight types of skin lesion images from the ISIC-2019 dataset

A. Image Pre-processing

Dermoscopic images in the ISIC 2019 dataset are of different sizes. Neural networks require input images of the same size, so all the images must be resized to a fixed dimension. Images with different dimensions such as 244 x 244, 250 x 250, 300 x 300, 310 x 310, 320 x 320, and 340 x 340 are tested. Among these, 300 x 300 gives the best result. So, images are resized into 300 x 300 dimensions.

B. Image Augmentation

Image augmentation technique significantly increases the diversity of the training set images [15]. In CNN, image augmentation techniques help to reduce the overfitting of the models, and it improves the performance of the training model [15]. An image augmentation technique provided by Keras is used for augmenting the images. This technique augments images when fetched rather than increasing the number of images beforehand. Different image augmentation techniques, such as horizontal and vertical flip, random rotation, random zoom, random contrast, and random height and width, are used to augment the image data. Image augmentation can be applied to the training set during fitting the model [16].

C. Transfer Learning

Transfer learning, a model re-training technique, is used in the proposed method because training a model from scratch

requires a large dataset. It is used on different pre-trained models, which are also fine-tuned. Fine-tuning involves unfreezing a few of the top layers of the frozen base model and re-training both newly added parts of the model and the top layers.

The number of images in the ISIC-2019 dataset is insufficient for training the Deep Learning (DL) models. So, models which are pre-trained on ImageNet are used, which have over 14 million images [17]. These pre-trained models are fine-tuned to get better accuracy. The authors of [18] have proved that transfer learning, and fine-tuning can improve classification accuracy. Other researchers have also shown that transfer learning can improve classification results [6], [7], [12], [19]. So, transfer learning is used on different pre-trained models. However, these models are initially used as feature extractors to avoid wrong predictions. Feature extraction means freezing all layers except the last fully connected layer and the output layer. The top-level classifier should be trained using a pre-trained model set to non-trainable before attempting fine-tuning. The magnitude of the gradient updates will be too big (due to the random weights from the classifier), and the pre-trained model will forget what it has learnt if a randomly initialized classifier is added on top of a pre-trained model. It is attempted to train all layers together [20]. A block diagram of the process of training a model using transfer learning is given in Figure 2.

IV. Dataset

The International Skin Imaging Collaboration (ISIC) 2019 dataset [21]–[23] is used in the proposed model to classify eight skin cancer types. This dataset is a skin image analysis challenge hosted by the International Skin Imaging Collaboration (ISIC). The ISIC-2019 challenge has only one goal, which is disease classification. This publicly available dataset has 25,331 images which are used for training across eight different categories, including Actinic Keratosis (AK), Basal Cell Carcinoma (BCC), Benign Keratosis (BKL), Dermatofibroma (DF), Melanoma (MEL), Melanocytic Nevus (NV), Squamous Cell Carcinoma (SCC), and Vascular lesion (VASC). Table I describes the image distribution of the dataset.

TABLE I: Dataset image distribution

SL No.	Class	No. of Images
1	AK	867
2	BCC	3,323
3	BKL	2,624
4	DF	239
5	MEL	4,522
6	NV	12,875
7	SCC	628
8	VASC	253

V. Evaluation and Testing

The classification result is evaluated using the standard metrics: accuracy, precision, recall, f1 score, and confusion matrix. Accuracy is not a good measure if the dataset is imbalanced. It can result in misleading interpretations. So, other classification measurements are required. Precision measures the number of true positives. On the other hand, recall measures the number of true negatives. F1 score provides a combined effect of precision and recall. The combined result of precision and recall helps to understand the quality of the model. F1 score is more useful than accuracy, especially if the dataset is imbalanced.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{1}$$

The confusion matrix gives better visibility of the classification results. For multi-class classification, a confusion matrix is necessary to demonstrate which classes the model confuses more while classifying.

VI. Model Training

The publicly available and labelled ISIC 2019 dataset has been used in this research. The dataset is categorized into eight classes, and the images are split into 80% training (20517 images), 10% test (2534 images), and 10% validation (2280 images) set.

Three additional dense layers, except the output layer, are added with the base pre-trained model. These three dense layers have 256, 128, and 64 neurons. Rectified linear activation unit (ReLu) is used in the dense layer as the activation function. L2 regularizer is used with these three dense layers to reduce overfitting. All the architectures in the pre-trained models consist of 1000 output categories. So, a dense layer with eight neurons is added to classify the input images into eight classes. The softmax activation function is used in the output layer. The architecture of the model is depicted in Figure 3.

During training the model, for the first ten epochs, the whole network weights are frozen except the last layer and used a learning rate of 0.001 with early stopping having the patience of 3 (i.e., the backpropagation algorithm will automatically terminate if there is no improvement in validation loss after three epochs). After that, the last few layers of the model are unfrozen and fine-tuned with a learning rate of 0.0001 for 100 epochs. This time early stopping having patience of 10, is used. If the validation accuracy is not increased for five epochs, the learning rate is decreased by a factor of 0.1. The adaptive momentum (Adam) optimizer is used for optimization, and the categorical cross-entropy loss function is used for updating the weights by back-propagating the error backwards.

VII. Implementation Details

Recently Kaggle, a subsidiary of Google, has become a popular online community platform among data scientists and machine learning enthusiasts. It provides virtual GPU memory to practitioners free of cost. So, the proposed model implementation is implemented in Kaggle, which provides a maximum of 13 GB RAM, 16 GB Graphics memory and 73 GB of storage space.

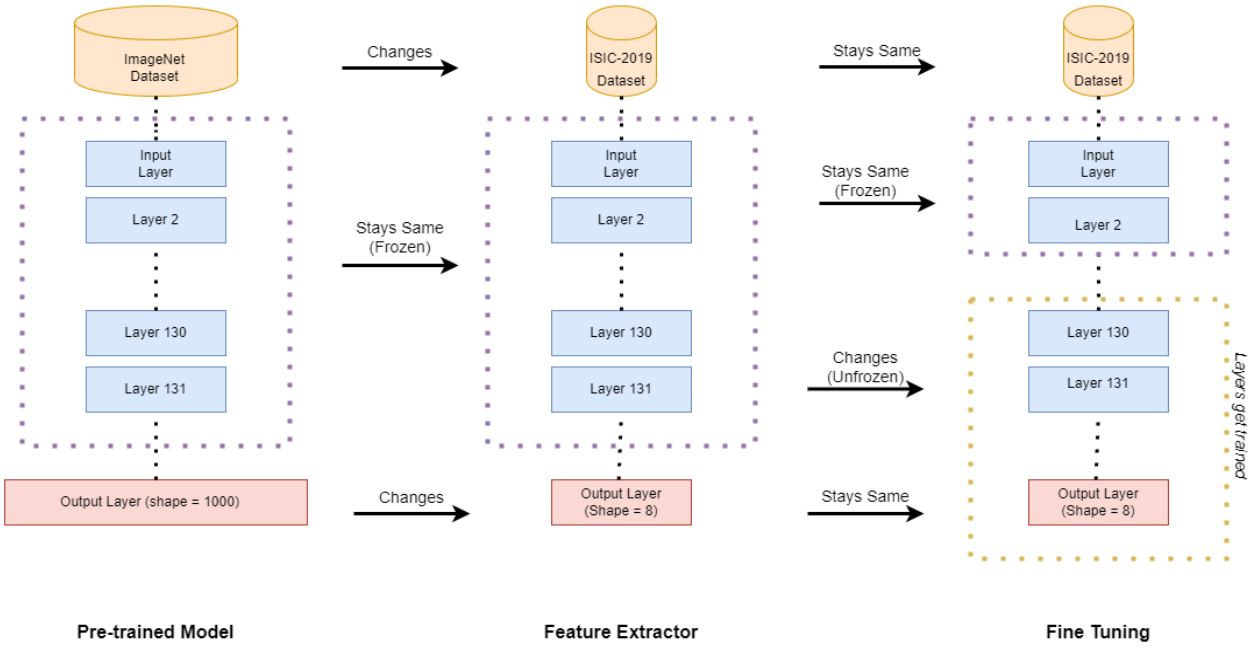

Fig. 2: Model training using transfer learning

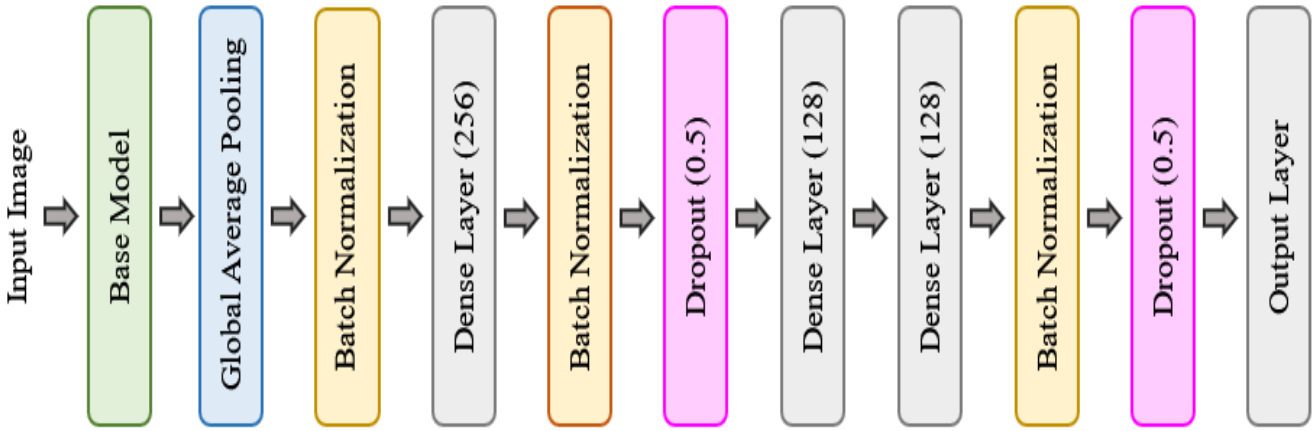

Fig. 3: Architecture of the implemented model

VIII. Experimental Results

In this paper, a model, Xception, is implemented, which is pre-trained on ImageNet. This section will discuss how the experimental result is obtained and evaluated. For the classification, the results were evaluated based on test accuracy, precision, recall, and f1-score. The proposed method has achieved 88% test accuracy and 99% training accuracy while [12] achieved 75.7% test accuracy by implementing Xception. In the proposed method, a larger dataset, hyperparameter tuning, adding three additional dense layers, and batch normalization have helped to improve test accuracy than [12]. Additionally, the proposed method has obtained the following:

The confusion matrix of the proposed method is depicted in Figure 5. The diagonal of the confusion matrix represents the true positive. As seen from the confusion matrix, the model could classify NV better than other classes. However, it is getting more confusing when classifying MEL and NV. For each class in a confusion matrix, True Positive (TP) is on its corresponding diagonal cell, False Positive (FP) is the summation of the values of corresponding rows except the TP value, False Negative (FN) is the summation of the values of corresponding columns except the TP value, The values of all columns and rows, excluding the values of the class for

	precision	recall	f1-score	support
0	0.77	0.71	0.74	87
1	0.92	0.90	0.91	332
2	0.81	0.79	0.80	263
3	0.94	0.62	0.75	24
4	0.84	0.80	0.82	452
5	0.92	0.96	0.94	1288
6	0.75	0.68	0.72	63
7	0.96	0.92	0.94	25
accuracy			0.88	2534
macro avg	0.86	0.80	0.83	2534
weighted avg	0.88	0.88	0.88	2534

Fig. 4: Performance of the implemented model

which we are computing the values, are added to create True Negative (TN).

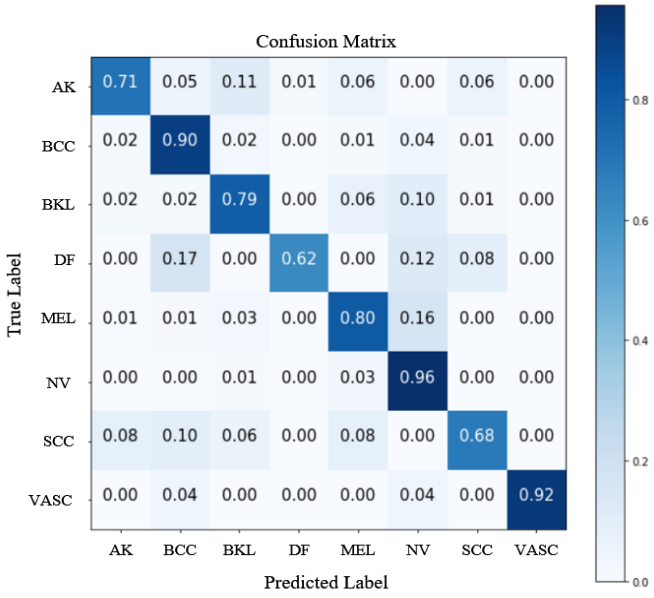

Fig. 5: Confusion matrix of the proposed method

A bar chart containing the accuracy compared with previous works has been depicted in Figure 6. Most authors have used classification accuracy as their evaluation method, but as only the classification accuracy is not a good measure, so, to evaluate classification accuracy in the proposed method, precision, recall, and f1 score are also used.

IX. Conclusion

Dermoscopic images can be used to detect skin cancer early, which is beneficial for both patients and medical professionals to speed up diagnosis and reduce costs. In this study, an efficient state-of-the-art pre-trained model, i.e., Xception, has been implemented for skin cancer classification. The model has tested on ISIC 2019 dataset and achieved 88% test

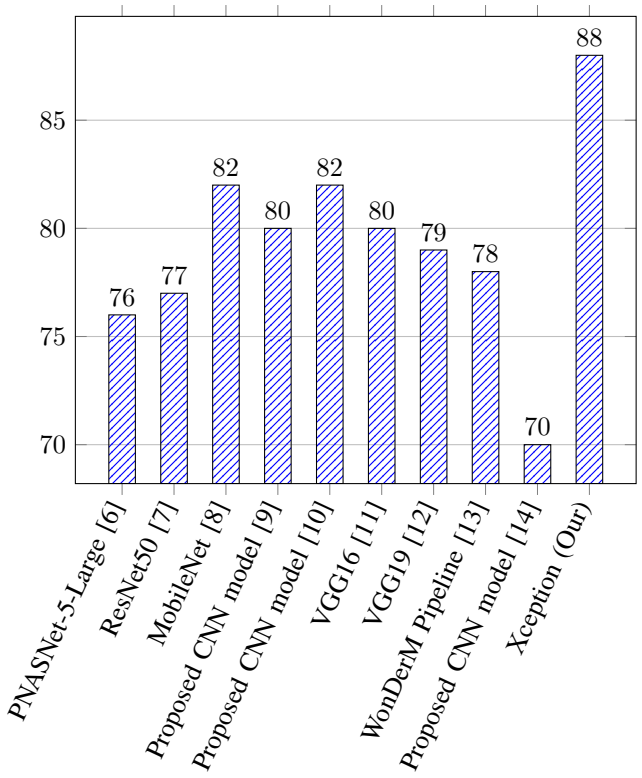

Fig. 6: Accuracy (in %) plot for different methods

accuracy, 88% f1 score, 88% precision, and 88% recall. By analyzing previous works, it can be said that this implemented state-of-the-art method has given a promising result, and it will be able to detect most skin lesions from dermoscopic images. Previous researchers have used comparatively fewer data to train their models. In this proposed method, data augmentation techniques have helped to improve classification accuracy, f1-score, precision and recall by increasing image data. This method will detect skin cancers early and reduce the chance of fatal skin damage. It will also be helpful for doctors to make quick decisions because people with darker skin colour are often diagnosed with skin cancer during their later stages when it becomes harder to treat. Future improvements may be possible by using a GPU with higher memory and carefully choosing the hyper-parameters. Especially, accuracy can be improved if the confusion between MEL and NV classes can be reduced as seen in Figure 5. A bar chart containing the accuracy compared with previous works has been depicted in Figure 6. Most of the authors have used classification accuracy as their evaluation criterion, but it solely is not a good measure and can mislead. So, to evaluate the proposed method, precision, recall, and f1 score are also provided.

References

[1] V. C. Hung, “4 ways you can get skin cancer besides sunlight.” [Online]. Available: https://www.drhungmd.com/blog/4-ways-you-can-get-skin-cancer-besides-sunlight/

[2] S. Serte and H. Demirel, "Gabor wavelet-based deep learning for skin lesion classification," *Computers in biology and medicine*, vol. 113, p. 103423, 2019.

[3] "American cancer society," 2022. [Online]. Available: https://cancerstatisticscenter.cancer.org

[4] C. H. O'Neill and C. R. Scoggins, "Melanoma," *Journal of Surgical Oncology*, vol. 120, no. 5, pp. 873–881, jun 2019.

[5] M. B. Lens and M. Dawes, "Global perspectives of contemporary epidemiological trends of cutaneous malignant melanoma." *British Journal of Dermatology*, 2004.

[6] M. A. A. Milton, "Automated skin lesion classification using ensemble of deep neural networks in isic 2018: Skin lesion analysis towards melanoma detection challenge," *arXiv preprint arXiv:1901.10802*, 2019.

[7] A. Pal, S. Ray, and U. Garain, "Skin disease identification from dermoscopy images using deep convolutional neural network," 2018.

[8] A. C. Salian, S. Vaze, P. Singh, G. N. Shaikh, S. Chapaneri, and D. Jayaswal, "Skin lesion classification using deep learning architectures," in *2020 3rd International conference on communication system, computing and IT applications (CSCITA)*. IEEE, 2020, pp. 168–173.

[9] S. R. S. Jianu, L. Ichim, and D. Popescu, "Automatic diagnosis of skin cancer using neural networks," in *2019 11th International Symposium on Advanced Topics in Electrical Engineering (ATEE)*. IEEE, 2019, pp. 1–4.

[10] L. Hoang, S.-H. Lee, E.-J. Lee, and K.-R. Kwon, "Multiclass skin lesion classification using a novel lightweight deep learning framework for smart healthcare," *Applied Sciences*, vol. 12, no. 5, p. 2677, 2022.

[11] T. Majtner, B. Bajić, S. Yildirim, J. Y. Hardeberg, J. Lindblad, and N. Sladoje, "Ensemble of convolutional neural networks for dermoscopic images classification," 2018.

[12] A. Aldwgeri and N. F. Abubacker, "Ensemble of deep convolutional neural network for skin lesion classification in dermoscopy images," in *International Visual Informatics Conference*. Springer, 2019, pp. 214–226.

[13] Y. C. Lee, S.-H. Jung, and H.-H. Won, "Wonderm: Skin lesion classification with fine-tuned neural networks," *arXiv preprint arXiv:1808.03426*, 2018.

[14] J. Rathod, V. Waghmode, A. Sodha, and P. Bhavathankar, "Diagnosis of skin diseases using convolutional neural networks," in *2018 second international conference on electronics, communication and aerospace technology (ICECA)*. IEEE, 2018, pp. 1048–1051.

[15] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, jan 2017.

[16] "Data augmentation in tensorflow tutorials," 2022. [Online]. Available: https://www.tensorflow.org/tutorials/images/data_augmentation

[17] "Devopedia," 2019. [Online]. Available: https://devopedia.org/imagenet

[18] R. Siddiqi, "Effectiveness of transfer learning and fine tuning in automated fruit image classification," in *Proceedings of the 2019 3rd International Conference on Deep Learning Technologies*, 2019, pp. 91–100.

[19] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[20] "Transfer learning and fine-tuning in tensorflow tutorials," 2022. [Online]. Available: https://www.tensorflow.org/tutorials/images/transfer_learning

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

[22] N. C. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler *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 *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*. IEEE, 2018, pp. 168–172.

[23] M. Combalia, N. C. Codella, V. Rotemberg, B. Helba, V. Vilaplana, O. Reiter, C. Carrera, A. Barreiro, A. C. Halpern, S. Puig *et al.*, "Bcn20000: Dermoscopic lesions in the wild," *arXiv preprint arXiv:1908.02288*, 2019.