Skin disease detection employing transfer learning approach-a fine-tune visual geometry group-19

Mohammad Al-Habib Islam, Sarkar Mohammad Shahriyar, Mohammad Jahangir Alam, Mushfiqur Rahman, Mohammad Rahmatul Kabir Rasel Sarker
Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh

Article Info

Article history:

Received Dec 31, 2022
Revised Feb 22, 2023
Accepted Feb 28, 2023

Keywords:

Convolutional neural network
Deep learning
Image processing
Skin disease
Transfer learning model

ABSTRACT

Your skin may become damaged by skin diseases and conditions. These illnesses can cause skin changes such as rashes, inflammation, itching, and other skin changes. While some skin conditions may run in families, others may result from a person's way of life. Skin conditions may be treated with pills, creams, ointments, changes in diet, and lifestyle modifications. Deep learning algorithms for computer vision applications have advanced quickly thanks to a significant amount of data for training the model and advancements in evaluation of proposed that can provide stronger simplifications. Undesired skin disease regions are eliminated, quality is raised, and the disease is tinted by discarding artifacts, decrease noise, and improving the image. Three augmentation techniques have raised the quantity of skin disease images. The five transfer learning models and various convolutional neural network (CNN) architectures analyzed the augmentation dataset. Visual geometry group-19 (VGG-19) offers the highest level of accuracy. Following the segmentation of the dermoscopic images, the affected skin cells' features are extracted using a feature extraction technique. The retrieved features are stratified using a CNN classifier, that is focused in deep learning. The best outcomes were obtained using the hyper-tuned VGG-19, which had test and validation accuracy of 99.21% and 99.25%, including both.

Corresponding Author:

Mohammad Jahangir Alam
Department of Computer Science and Engineering, Daffodil International University
Dhaka, Bangladesh
Email: jahangir.cse@diu.edu.bd

1. INTRODUCTION

One form of cancer that begins in skin cells and can infiltrate or spread to other bodily areas is skin cancer. The proliferation of aberrant cells is the primary cause of skin cancer. In 2020, it was ranked as the fourth most prevalent cancer. For a variety of reasons, determining the prevalence of skin cancer is a difficult task. Skin cancer comes in a variety of forms, and cancer registries frequently ignore non-melanoma skin cancer [1]. Since the most prevalent forms of skin cancer treatment are surgery or ablation, registrations are often lacking. Because to these factors, the occurrence of skin cancer globally is likely underestimated. The Scandinavian nations in Europe are next in order of reported skin cancer rates, with New Zealand and Australia already having highest rates (33.6 per 100,000 and 33.3 per 100,000 respectively). The obvious cause of this high prevalence is sun exposure, which is what causes the majority of skin cancers. The human head and neck are the areas where skin malignancies are most often discovered [2]. According to the global cancer observatory, this kind of cancer is less prevalent in Southeast Asia [3], [4]. Skin cancer may be inspected by a dermatologist early on since it manifests on the exterior of the body and is a visible sort of sickness. Early detection of skin cancer lesions lowers morbidity, decreases the cost of healthcare, and increases patient survival rates. The incidence of UV radiation brought on by exposure to sunshine is directly correlated with the existence of skin cancer [5].

This study proposes using a fine-tuned and hyper-tuned version of the visual geometry group-19 (VGG-19) deep learning model, which has been built on transfer learning and subjected to an ablation study, as a fully automated and efficient solution for categorizing photos of various skin conditions. The purpose of this research is to improve patient outcomes and lower the risk of death by early identification and classification of skin conditions. This is achieved by incorporating specialists who can provide more effective and efficient treatment. For a convolutional neural network (CNN) model to implement perfectly, noise and artifact removal is essential and were performed on the dataset. Additionally, the similarity between disease regions and impenetrable skin tissue may make interpretation difficult. Raw image brightness and contrast levels were balanced to make disease lesions more visible.

When tested on the PHDB Melanoma dataset, the model suggested in this study showed an improvement over the base method in accuracy as well as computational efficiency [6]. Additionally, a skin cancer app for iOS devices was developed by Dai *et al.* [7] using CNN. The trained model was developed using data from the HAM10000 [8]. The PAD-UFES-20 dataset now includes more skin lesions [9]. They were able to help their network grow more sophisticated and discriminative traits, which ultimately improved performance. They used extremely deep residual networks for classification and fully convolutional residual networks (FCRN) for segmenting skin lesions in their two-stage framework. The 85.5% was the highest reported classification accuracy for this challenge.

Tschandl *et al.* [8] By using ISBI 2016 challenge datasets, CNN with more than 50 layers was developed in 2016 again for detection of malignant melanoma cancer. Their approach was distinctive in that it produced a classification accuracy of 85.5% while using far less training data than other authors and a deeper network. Convolutional neural networks (CNNs) were used by Estevan *et al.* [10] to categorize skin lesions. Then, using clinical photos that had been verified by a biopsy, they put it to the test against 21 board-certified dermatologists. It has been shown that artificial intelligence systems may be able to identify skin cancer with expertise on par with that of dermatologists. The accuracy for carcinoma pictures was 96%, melanoma images were 94%, and thermoscopic images of melanoma were 91%.

Jinnai *et al.* [11] utilized deep learning to develop a method for categorizing pigmented skin lesions into skin cancer. Reports state that clinical diagnosis accuracy for melanoma detection ranges from 65% to 80% [12]. When dermoscopic images are used, the accuracy of diagnosing skin lesions is increased by 49% [13]. This study effectively compares a novel deep learning algorithm for classifying skin lesions against a collection of dermoscopic skin lesion images (the international skin imaging collaboration (ISIC) archive dataset [14]). It obtained an accuracy value of 81.33% [15] when utilizing the dataset partitioned exactly as the ISBI 2016 Challenge, which would position the suggested technique in the top three in that task.

In 2020, Ashraf *et al.* [16] presented a framework for skin cancer diagnosis that utilizes transfer learning and an intelligent region of interest (ROI) technique. The ROI-based strategy achieved estimation accuracy of 97.9% on their initial data and 97.4% on their other new dataset, outperforming earlier systems that classified images based on their entirety (global features). To complement data, Goyal and Yap [17] suggested employing the ROI identification on necroscopic pictures in 2018.

In 2020, Ali *et al.* [18] unveiled a distinctive fuzzy multilayer perceptron (F-MLP) technique to aid in the early diagnosis of melanomas. This system aims to identify irregularity in the perimeter of skin lesions. Additionally, in the middle of 2019 Fujisawa *et al.* [19] published an autonomous skin cancer categorization algorithm that utilizes deep learning.

The ILSVR2012 database, that includes of 1.2 million pictures divided into 1,000 classes was used to train this system. By merging convolutional neural networks and textural characteristics, Alizadeh and Mahloojifar [20] created an automated skin cancer diagnosis method using thermoscopic pictures. They utilized the datasets PH2, ISIC 2016, and ISIC 2019. For three datasets, this automated approach achieved accuracy of 85.2%, 96.7%, and 97.5%, respectively.

2. METHOD

This study investigated five models to determine the most accurate network and the best transfer learning method for the problem of identifying skin conditions. These five transfer learning algorithms VGG-16, VGG-19, MobileNetV2, InceptionV3, and MobileNet-have all been developed and tested using test data.VGG-16 is one of the best models of transfer learning approaches. The deep convolutional neural network (DCNN) model recognized as VGG16 was first presented by Simonyan and Zisserman [21]. Each source channel receives a different filter thanks to the depthwise convolution technique used by the MobileNets [22]. With the exception of using inverted residual blocks with bottlenecking features and eliminating nonlinearities in narrow layers, MobileNetV2 architecture is very similar to the original MobileNet [23].

The flow diagram in Figure 1 begins by taking the dataset and applying several image pre-processing techniques to process the images. Data augmentation is then used to balance the dataset. The balanced dataset is then run through five transfer learning models, with VGG-19 emerging as the best model. The model is then fine-tuned for greater accuracy using an ablation study. Finally, the results are analyzed to assess the performance of the model.

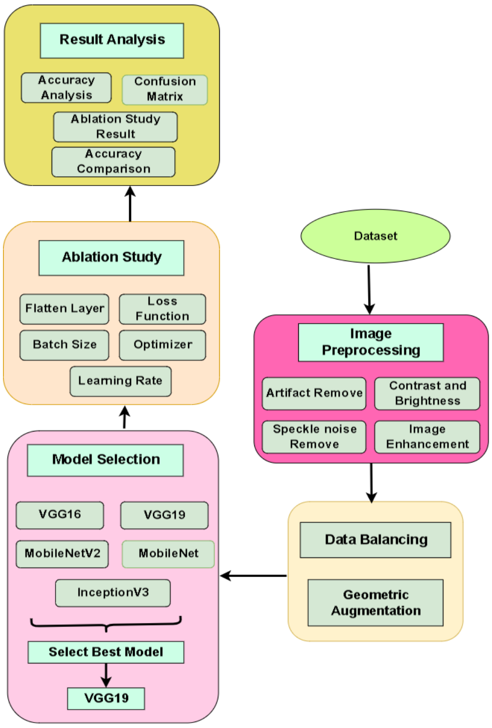

Figure 1. Workflow of full categorization system

2.1. Ablation study

The addition of numerous concepts frequently improves the performance of the final model when building a special transfer learning model. To understand the impact of each of these innovations separately in a study, though, is helpful. Researchers frequently evaluate their models with each component disabled in order to gauge the effects of individual components, quantifying the efficiency loss entire method.

2.2. Dataset description and training approach

The open-source website Kaggle served as the source of the data for this research. A total of 2,357 images from the dataset were examined for this study. The dataset used in this study encompasses nine classes. The images in the dataset have different size of pixels in grayscale system. To make the final prediction, 20% of the dataset was used as the test dataset in a recent study. With a ratio of 70:20:10, the images were split into three sets: train, validation, and test. The batch size for training the models is 16, and the highest epoch number is 100 [24]. At the time of training the model, callbacks can be used to get a view of the model's internal states and statistics. When you want to automate some tasks that give you control over the training process after each training/epoch, you define and use a callback. The Adam optimizer often outperforms all other optimization methods, requires less tuning, and has a faster computation time.

2.3. Image preprocessing

This research focuses on image processing to improve the model's performance. Because the skin disease dataset photos have a lot of noise and artifacts. Image processing is frequently thought of as the arbitrary manipulation of an image to meet an aesthetic requirement or to support the desired reality. Processing images is the initial stage in building a deep learning algorithm because they contain many artifacts.

Figure 2 depicts image processing techniques applied to the skin cancer dataset. Morphological dilation techniques are used to remove artifacts from the skin cancer dataset, while the dataset's speckle noise is removed using median filtering. Clahe is used to enhance low-contrast images by restricting contrast and implementing a clipping limit to address noise amplification. The images in the study are adjusted using alpha and beta correction to find the desired images. Median filtering is important because it maintains edges while removing noise.

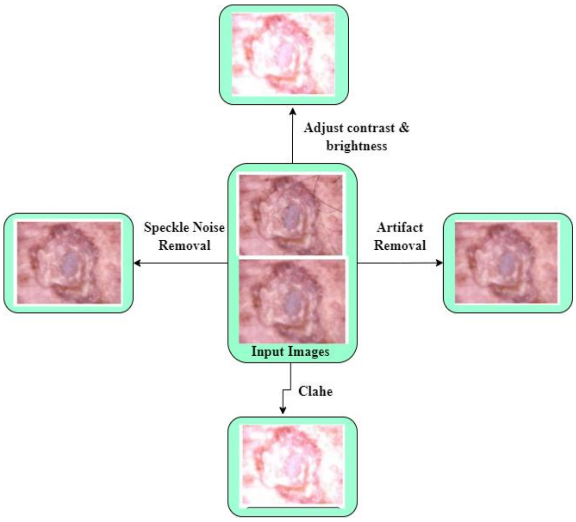

Figure 2. Pre-processing of the image

2.4. Verification

The most basic and widely used loss function is undoubtedly mean squared error (MSE) [25]. It shows the squared cumulative error between the uncompressed and compressed versions of the image. The error is inversely correlated with MSE value. As peak signal-to-noise ratio (PSNR) increases, the quality of the compressed or reconstructed image also increases. Preprocessing algorithms reduce image quality, as measured by structural similarity index measure (SSIM).

3. RESULTS AND DISCUSSION

3.1. Result of transfer learning model

Table 1 provides insight into the performance of different transfer learning models on skin disease classification task. The models were evaluated based on their training, testing, and validation accuracy, training, testing, and validation loss. Among the models, VGG-19 performed the best with the highest scores in all three accuracy categories, while inceptionV3 had the lowest accuracy score.

Table 1. Five transfer learning model accuracy

Model	Training accuracy	Testing accuracy	Val accuracy	Training loss	Testing loss	Val loss
VGG-16	95.98	95.65	95.66	0.33	0.29	0.31
VGG-19	96.29	96.16	96.16	0.2	0.16	0.15
Mobile Net	94.66	94.35	94.35	0.24	0.26	0.26
Mobile net V2	94.48	94.26	94.27	0.28	0.35	0.34
InceptionV3	89.75	88.61	88.61	0.36	0.39	0.39

3.2. Result of ablation study

A process of improving classification accuracy and reliability by making some design changes. An optimized VGG-19 architecture was used to modify various design elements in five separate experiments, which collectively make up the ablation study. By systematically altering and testing different design elements, the the most effective configurations for improving the accuracy and reliability of the model was identified.

3.2.1. Changing flatten layer

As seen in Figure 3, by using flatten layer provides the greatest accuracy. The global maximum pooling and global average pooling both provide poor accuracy. When the layer is flattened, its accuracy is 96.29%; when it is global maximum pooling, it is 92.65%; and when it is global average pooling, it is 90.52%.

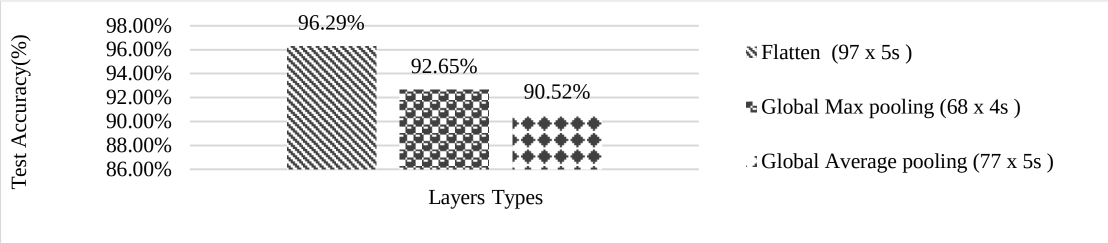

Figure 3. Changing flatten layer

3.2.2. Changing the batch size

When batch size 32 was used, we found the highest accuracy, as shown in Figure 4. Furthermore, batch size 16 and batch size 64 offer poor accuracy. The accuracy is 97.16% when the batch size is 32, 94.55% when the batch size is 64, and 96.29% when the batch size is 16.

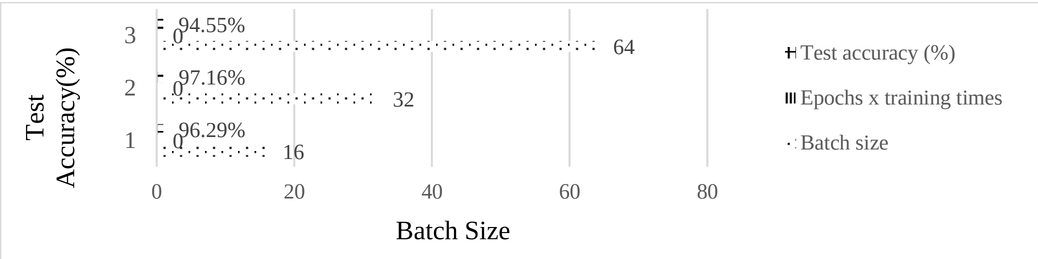

Figure 4. Change the batch size

3.2.3. Changing the loss function

Loss functions quantify the difference between a model's predicted and actual output. Changing the loss functions may affect how well the model performs. Figure 5 shows that when categorical cross-entropy was utilized as a loss function, the accuracy that was obtained under this specific instance was the highest, 97.16%.

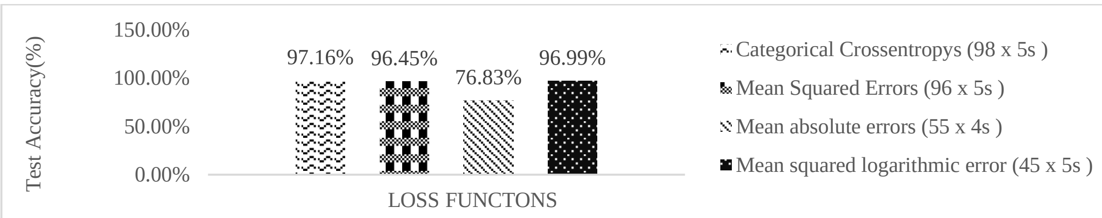

Figure 5. Changing losses function

3.2.4. Changing optimizer

The results depicted in Figure 6 illustrate that when compared to other optimizers like Nadam and Adamax, the Adam optimizer offers the highest accuracy. This suggests that the Adam optimizer is better suited for the given task than the other optimizers tested. Therefore, it is recommended to use the Adam optimizer for this particular task.

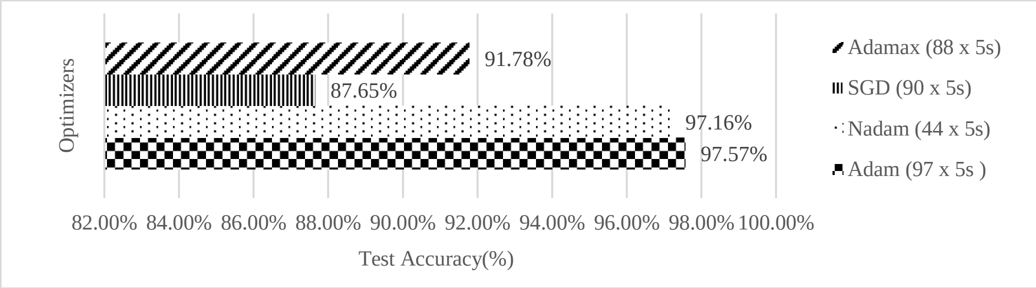

Figure 6. Changing the optimizer

3.2.5. Changing learning rate

The findings in Figure 7 indicate that, in comparison to 0.0001, 0.01, and 0.1 learning rates, a learning rate that is 0.001 is the model's most productive rate. This implies that the learning rate is a critical hyperparameter in deep learning, and using the appropriate learning rate can result in higher accuracy. The results also indicate that lower learning rates result in a slower convergence, while higher learning rates may cause instability or divergence during training.

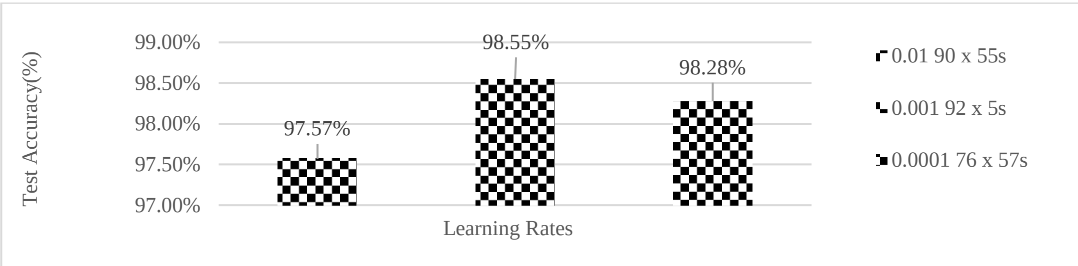

Figure 7. Changing learning rate

3.2.6. Performance analysis of best model

The configuration of the proposed model in this research paper was optimized for best results. The image sizes used in the model were 224×224 pixels. As for optimization Adam was utilized where the algorithm was taught over a period of 90 epochs.32 was batch size and the learning rate was tuned to 0.001. The activation function used in the model was Softmax and the dropout rate was 0.5. The momentum was set to 0.9 and the accuracy achieved by the model was 98.55%.

3.2.7. Statistics and performence evaluation

Table 2 presents the evaluation metrics for the best-performing hyper-tuned VGG-19 model, including the precision, accuracy, specificity, recall, F1 score, false discovery rate (FDR), false negative rate (FNR), false positive rate (FPR), and root mean square error (RMSE). These measures are frequently employed in deep learning to assess how well categorization models perform. The values reported in Table 2 provide insight into the effectiveness of the VGG-19 model and its ability to accurately classify images.

Table 2. Statistics and performance evaluation

Accuracy	FPR (%)	FDR (%)	FNR (%)	RMAE	MSE	Precession	Recall	Specifity	F1 Score
98.55	1.68	2.58	2.51	5.61	2.23	96.98	96.98	96.19	96.11

4. CONCLUSION

To categorize skin diseases, this paper presents an approach that makes use of CNNs and transfer learning techniques. A method has been developed that may assist both people and medical professionals in detecting skin cancer classes, regardless of whether the cancer is benign or malignant. A conclusion that can be drawn from the experimental and assessment phase is that the model may be regarded as a standard for diagnosing skin diseases by providing assistance to medical experts. Any physician may recognize the proper findings by capturing a few random photos, but the old technique takes too much time to determine whether or not a case has been correctly identified. Transfer learning models performed significantly better for multiclass classification in this study than traditional classifiers. Despite the research's serious flaw-the loss of a sizable amount of accurate medical data-the dataset for the proposed model is not large enough. The clinical use of deep learning for the diagnosis of more disorders can be investigated in future studies. For relatively uncommon diseases, transfer learning may be helpful. Additionally, models might develop to require fewer preprocessing steps. Along with these, a deeper comprehension of the reconstruction kernel or image thickness might enhance the performance of deep learning methods.

REFERENCES

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, Nov. 2018, doi: 10.3322/caac.21492.

[2] Cancer Research UK, "Prostate cancer incidence statistics," *Cancer Research UK*, 2013. http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer/incidence (accessed Mar. 02, 2022).

[3] WHO, "The global cancer observatory," *International Agency for Research on Cancer*, 2021. https://www.who.int/health-topics/cancer#tab=tab_1 (accessed Dec. 06, 2021).

[4] Macro Trends, "World life expectancy," *Macro Trends*, 2022. https://www.macrotrends.net/countries/WLD/world/life-expectancy (accessed Nov. 22,2022).

[5] WHO, "WHO | Health effects of UV radiation," 2013. http://www.who.int/uv/health/en/ (accessed Jul. 26, 2022).

[6] M. Berseth, "ISIC 2017-skinlesion analysis towards melanoma detection," 2017, *arxiv: 1703.00523*.

[7] X. Dai, I. Spasic, B. Meyer, S. Chapman, and F. Andres, "Machine learning on mobile: An on-device inference app for skin cancer detection," in *2019 4th International Conference on Fog and Mobile Edge Computing, FMEC 2019*, Jun. 2019, pp. 301–305, doi: 10.1109/FMEC.2019.8795362.

[8] P. Tschandl, C. Rosendahl, and H. Kittler, "Data descriptor: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data*, vol. 5, no. 1, p. 180161, Aug. 2018, doi: 10.1038/sdata.2018.161.

[9] A. G. C. Pacheco *et al.*, "PAD-UFES-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones," *Data in Brief*, vol. 32, p. 106221, Oct. 2020, doi: 10.1016/j.dib.2020.106221.

[10] A. Esteva *et al.*, "Corrigendum: Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 546, no. 7660, p. 686, Jun. 2017, doi: 10.1038/nature22985.

[11] S. Jinnai, N. Yamazaki, Y. Hirano, Y. Sugawara, Y. Ohe, and R. Hamamoto, "The development of a skin cancer classification system for pigmented skin lesions using deep learning," *Biomolecules*, vol. 10, no. 8, pp. 1–13, Jul. 2020, doi: 10.3390/biom10081123.

[12] G. Argenziano and H. P. Soyer, "Dermoscopy of pigmented skin lesions-a valuable tool for early diagnosis of melanoma," *Lancet Oncology*, vol. 2, no. 7, pp. 443–449, Jul. 2001, doi: 10.1016/S1470-2045(00)00422-8.

[13] H. Kittler, H. Pehamberger, K. Wolff, and M. Binder, "Diagnostic accuracy of dermoscopy," *Lancet Oncology*, vol. 3, no. 3, pp. 159–165, Mar. 2002, doi: 10.1016/S1470-2045(02)00679-4.

[14] International Skin Imaging Collaboration (ISIC), "International skin imaging collaboration: Melanoma project," 2010. https://www.isic-archive.com/ (accessed Oct. 22, 2022).

[15] IEEE, "IEEE-ISBI 2023-International symposium on biomedical imaging," *Biomedicalimaging.org*, 2023. https://2023.biomedicalimaging.org/en/ (accessed Mar. 01, 2023).

[16] R. Ashraf *et al.*, "Region-of-interest based transfer learning assisted framework for skin cancer detection," *IEEE Access*, vol. 8, pp. 147858–147871, 2020, doi: 10.1109/ACCESS.2020.3014701.

[17] M. Goyal and M. H. Yap, "Region of interest detection in dermoscopic images for natural data-augmentation," 2018, *arXiv: 1807.10711*.

[18] A. R. Ali, J. Li, S. Kanwal, G. Yang, A. Hussain, and S. Jane O'Shea, "A novel fuzzy multilayer perceptron (F-MLP) for the detection of irregularity in skin lesion border using dermoscopic images," *Frontiers in Medicine*, vol. 7, Jul. 2020, doi: 10.3389/fmed.2020.00297.

[19] Y. Fujisawa, S. Inoue, and Y. Nakamura, "The possibility of deep learning-based, computer-aided skin tumor classifiers," *Frontiers in Medicine*, vol. 6, Aug. 2019, doi: 10.3389/fmed.2019.00191.

[20] S. M. Alizadeh and A. Mahloojifar, "Automatic skin cancer detection in dermoscopy images by combining convolutional neural networks and texture features," *International Journal of Imaging Systems and Technology*, vol. 31, no. 2, pp. 695–707, Jun. 2021, doi: 10.1002/ima.22490.

[21] P. Wang, J. Wang, Y. Li, P. Li, L. Li, and M. Jiang, "Automatic classification of breast cancer histopathological images based on deep feature fusion and enhanced routing," *Biomedical Signal Processing and Control*, vol. 65, p. 102341, Mar. 2021, doi: 10.1016/j.bspc.2020.102341.

[22] A. G. Howard *et al.*, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv: 1704.04861*.

[23] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 4510–4520, doi: 10.1109/CVPR.2018.00474.

[24] Shallu and R. Mehra, "Breast cancer histology images classification: Training from scratch or transfer learning?," *ICT Express*, vol. 4, no. 4, pp. 247–254, Dec. 2018, doi: 10.1016/j.icte.2018.10.007.

[25] I. U. Khan *et al.*, "An effective approach to address processing time and computational complexity employing modified CCT for lung disease classification," *Intelligent Systems with Applications*, vol. 16, p. 200147, Nov. 2022, doi: 10.1016/j.iswa.2022.200147.

BIOGRAPHIES OF AUTHORS

Mohammad Al-Habib Islam is currently a final year student at Daffodil International University, Department of Computer Science and Engineering. He has a passion for exploring various areas of research, including machine learning, deep learning, and image processing. He is reachable via email: habib-15-2629@diu.aedu.bd.

Sarkar Mohammad Shahriyar is currently pursuing his final year of study in The Computer Science and Engineering Department at Daffodil International University. His areas of research focus include machine learning, deep learning, and IoT. He is reachable via email: shahriyar15-2667@diu.edu.bd.

Mohammad Jahangir Alam has been a senior lecturer at Daffodil International University since 2017. He is a consistent research enthusiast interested in embedded systems, IoT, machine learning and artificial intelligence. He has now been part of numerous efficient research publications as author/co-author. Along with research, he has expertise in software application development. He is reachable via email: jahangir.cse@diu.edu.bd.

Mushfiqur Rahman is a senior faculty member in the department of CSE at Daffodil International University. He is a research enthusiast with interests in artificial intelligence, machine learning, embedded systems, and the internet of things. As an author or co-author, he has now contributed to many fruitful research articles. He has experience in research and software application development. He is reachable via email: mushfiqur.cse@diu.edu.bd.

Mohammad Rahmatul Kabir Rasel Sarker has been a faculty member at Daffodil International University since 2019. He is a consistent research enthusiast interested in deep learning, machine learning, human computer interaction. He has now been part of numerous efficient research publications as author/co-author. He is reachable via email: raselsarker.cse@diu.edu.bd.