



Enhancing Skin Disease Diagnosis Through Fine-Tune Convolutional Neural Network: A Comparative Study with Multi-class Approach

Najnin Akter Ringky, Abu Kowshir Bitto*, Khalid Been
Md. Badruzzaman Biplob, Md. Fazla Elahe, Musabbir Hasan Sammak &
Tapushe Rabaya Toma

Department of Software Engineering, Faculty of Science and Technology,
Daffodil International University, Daffodil Smart City, Birulia, Savar, Dhaka-1216,
Bangladesh

*E-mail: abu.kowshir777@gmail.com

Abstract. Due to their similar appearance, skin disorders frequently disguise their early warning signs from our skin, which is the defense system of the body. Preventing serious disorders requires their early detection. This work investigated the use of fine-tune transfer learning as a fast and accurate way to diagnose skin diseases. To classify different skin issues, we used pre-trained models, i.e., InceptionV3, DenseNet201, and Xception. This work examined 17,500 photos from three sources. It was found that fine-tune Xception performed exceptionally well, with an accuracy rate of 99.14%. It was closely followed by DenseNet201 and InceptionV3, each with different processing speeds, 98.74% and 98.46%, respectively. We used transfer learning with data sets validated by medical experts, outperforming earlier research in precision. This more accurate detection of skin diseases could greatly improve patient outcomes and expedite medical procedures. This approach is new in that it fine-tunes transfer learning by utilizing a vast number of data to increase accuracy compared to other researcher works.

Keywords: *fine-tune; imbalance data; medical imaging; skin disease; Xception.*

1 Introduction

Skin diseases have a profound global impact, affecting around 900 million people worldwide with over 3,000 different skin-related conditions [1]. As the body's largest organ, the skin serves multiple functions, including protection, nutrient absorption, waste disposal, temperature regulation, and sensation perception. The burden of skin diseases extends beyond personal well-being, putting significant financial pressure on healthcare systems, covering a spectrum from common issues like acne to life-threatening diseases like melanoma.

Rapid detection of skin diseases is crucial to halt their progression, reduce healthcare costs, and enhance patients quality of life. The objective of this study

was to improve the precision of skin disease diagnosis by several different data sets and cutting-edge methodologies.

Recent advancements in machine learning and artificial intelligence have revolutionized medical diagnostics [2]. Transfer learning, a prominent technique in machine learning, offers a promising approach to complex medical diagnostic challenges, even with limited data. This study aimed to enhance skin disease detection precision by exploring transfer learning strategies, adapting pre-trained models initially designed for tasks like image identification. Employing a fine-tune convolutional multi-layer feature network, this research addressed the limitations of single-source data and manual diagnosis, using a fine-tune CNN model for higher accuracy. The primary research question focused on the need for an effective and precise approach to skin disorder detection, utilizing various data sources. The goal was to create and apply artificial intelligence and machine learning algorithms for automated skin disease diagnosis, improving accuracy without the need for large, specialized data sets.

2 Literature Review

Numerous studies have delved into skin disease diagnosis using various machine learning and deep learning techniques. Hosny *et al.* [3] used AlexNet, surpassing existing skin cancer classification methods with 96.82% accuracy. El Gannour *et al.* [4] employed ensemble learning, achieving 96.03% accuracy using the ISIC 2018 data set.

Domain-specific augmentation approaches were explored by Jahanifar *et al.* [5], enhancing model generalization. Bakkouri *et al.* [6] used Convolutional Fusion Unit (CFU) to improve skin abnormality classification. Abunadi *et al.* [7] used hybrid features and neural networks, achieving 97.91% accuracy for the PH2 data set. Shaheen *et al.* [8] applied particle swarm optimization and pre-trained CNNs for skin disease classification, achieving 97.82% accuracy. Raghavendra *et al.* [9] proposed a DCNN model with a graphical user interface, achieving 97.20% accuracy. Thanka [10] achieved 99.1% accuracy using ISIC data sets with VGG16 + XGBOOST. Alani *et al.* [11] presented a multiclass skin lesion diagnosis system, achieving 95% accuracy. Adegun *et al.* [12] combined segmentation and classification with an FCN-based DenseNet, obtaining 98% accuracy. Mehmood [13] introduced SBXception with 96.97% accuracy on the HAM10000 data set. Goyal *et al.* [14] used deep learning for multi-class segmentation, achieving 84.62% accuracy for melanoma diagnosis. Balaji *et al.* [15] employed dynamic graph cut and Naïve Bayes, achieving 92.8% accuracy. Al-Masni *et al.* [16] integrated skin disease classification and segmentation, achieving 89.28% accuracy with ResNet-50. Sonawane *et al.* [17] developed a web-based application with DenseNet-121, achieving 84.64% accuracy. Usama

et al. [18] proposed an automated approach for skin lesion segmentation, achieving 95.9% accuracy. Gairola *et al.* [19] proposed a fully fused network (FFN), which consists of an improved fusion block (IFB) and an improved single block (ISB). The greatest accuracy utilizing HAM10000 (for ResNet101V2), HAM10000 (for Resnet50 + ResNet101V2), and HAM10000 (for Resnet50 + ResNet101V2) was 86% for ISB, 90% for IFB, and 92% for FFN in the experiments. Purni *et al.* [20] employed the ISIC 2018 and 2019 public data sets to identify and detect multi-class skin cancer using an enhanced deep learning technique called convolution neural network (CNN) for classification and an improved canny edge for detection, with an accuracy of 99%.

These studies showcase the diversity of approaches in skin disease diagnosis, using advanced technologies to achieve high accuracies across various data sets. The challenges of data imbalance, model interpretability, and ethical considerations remain essential areas for future research. We used diverse data set for large scale data set with fine-tune transfer learning to gain more accuracy compared to other research with a balanced data set and with ethical considerations.

3 Methodology

The stepwise approach to achieving our goal involved various phases, including data set collection, data preprocessing, splitting the data set for training and testing then model development, with model fine-tuning, performance selection and best model selection. Figure 1 illustrates the operational procedure.

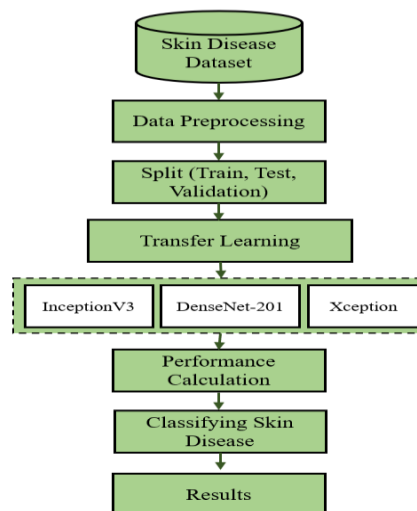


Figure 1 Step-by-step representation of our working procedure.

3.1 Data Set

We gathered data from Kaggle [21] for 10 classes, each having 1,250 data points (Eczema, Melanoma, Atopic Dermatitis, Basal Cell Carcinoma, Melanocytic Nevi, Benign Keratosis-like Lesions, Psoriasis, Lichen Planus, and related diseases, Seborrheic Keratoses and other Benign Tumors, Tinea Ringworm Candidiasis, and other Fungal Infections, Warts Molluscum, and other Viral Infections). Additionally, we collected 480 data points of Squamous cell carcinoma, 650 data points of Acne, and 344 data points of Dermatofibroma from HARVARD DATAVERSE [22]. Furthermore, 570 data points of Vascular lesion were obtained from STANFORD AIMI [23]. We combined these to create a single data set with medical verification from Dr. Mohammad Kykobad Hosain, MBBS, MCPS, DMF (BSMMU), CCD (BIRDEM), Managing Director, Sahara Modern Hospital & Diagnostic Centre, Bangladesh.

3.2 Data Preprocessing

To enhance the effectiveness of image classification, several crucial preprocessing steps are employed on the images in the proposed method [24]. These include rotation, scaling, and translation. The images are scaled to a 448 x 448 size for Inception V3, DenseNet201, and Xception, with a batch size of 64 and a learning rate of 0.01 for each model, ensuring compliance with specified requirements while maintaining photo quality. The pixel value is normalized between -1 and +1. Augmentation techniques such as rotation, width shift, height shift, shear range, and horizontal flip are employed to enhance the images.

While 10 classes had sufficient images (1,250), the remaining 4 classes did not (480 data of Squamous cell carcinoma, 650 data of Acne, 344 data of Dermatofibroma, 570 of vascular lesion). Augmentation was used to balance all classes to 1,250 images for each of the 4 classes. We used 14,000 images for training, 1,750 images for testing, and 1,750 for validation.

3.3 Performance Calculation

Several image classification performance evaluation metrics have already been evaluated by other researchers [25,26]. We employed Equations 1-4 to estimate the effectiveness of the classifiers in identifying skin conditions. In the formulas below, TP represents true positive, FP is false positive, TN is true negative, and FN is false negative.

$$Accuracy = \frac{True\ Positive + True\ Negative}{Total\ Number\ of\ Images} \times 100\% \quad (1)$$

$$Precision = \left(\frac{TP}{TP+FP} \right) \times 100\% \quad (2)$$

$$F1\ Score = (2 \times \frac{Precision \times Recall}{Precision + Recall}) \times 100\% \quad (3)$$

$$Recall = \frac{TP}{(TP + FN)} \times 100\% \quad (4)$$

4 Analysis and Discussion

The experiment was conducted on 17,500 images related to skin diseases. The images were categorized into 80% for training, 10% for testing and 10% for validation, i.e., 14,000 training images, 1,750 test images, and 1,750 validation images. In the whole model, the learning rate during training was modified (lr = 0.01) based on training and validation accuracy using a custom callback class called LRA. The models were constructed using categorical cross-entropy loss and the Adamax optimizer.

For learning rate changes, the LRA callback was utilized after a predetermined 50 epochs of training. Figure 2 shows plots of the training and validation loss (best number of epochs = 50) and the training and validation accuracy (best number of epochs = 46) over the number of epochs, executed by the script after training for Inception V3. The model's final few layers were unfrozen to allow for another 20 epochs of fine-tuning. Figure 3 shows that, after fine-tuning over 70 epochs, the best number of epochs was 70 for training and validation loss and the best number of epochs was 51 for training and validation accuracy. Model metrics after training gave a loss of 0.25455 and an accuracy of 0.98457. A function called print_info was defined in the script to print details concerning model predictions, such as incorrectly categorized images and class-wise errors shown in Figure 4. Table 1 shows the class wise performance on the Inception V3 model.

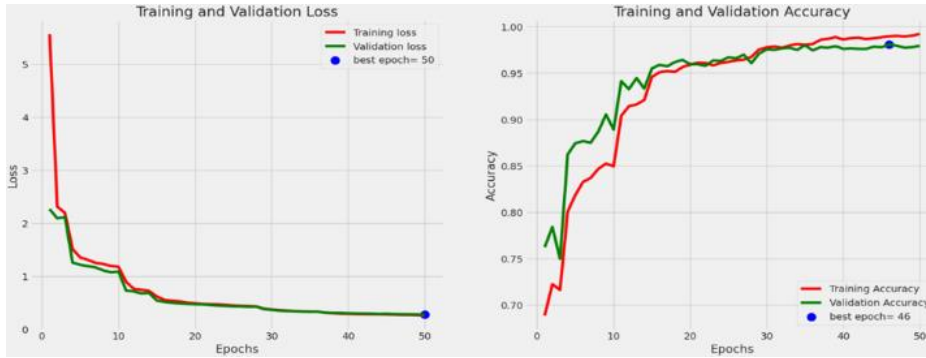


Figure 2 Training and validation loss and accuracy over 50 epochs for Inception V3.

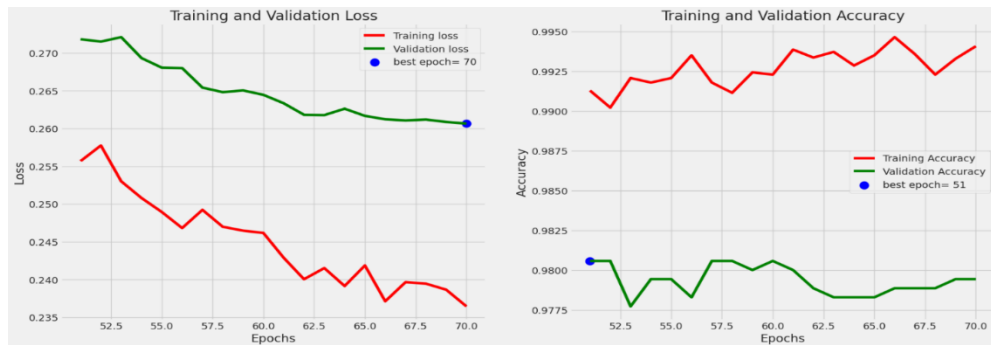


Figure 3 Fine-tune training and validation loss and accuracy over 70 epochs for Inception V3.

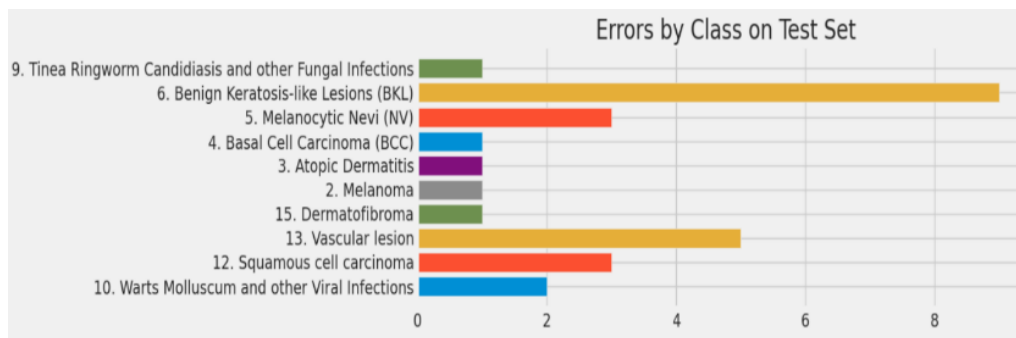


Figure 4 Errors by class set on Inception V3.

Table 1 Class wise performance on Inception V3.

Classes	Precision	Recall	F1	Support
Eczema	1.00	1.00	1.00	129
Warts molluscum and other viral infections	0.99	0.98	0.99	114
Acne	1.00	1.00	1.00	131
Squamous cell carcinoma	0.98	0.98	0.98	121
Vascular lesion	0.96	0.96	0.96	112
Dermatofibroma	0.96	0.99	0.98	133
Melanoma	0.96	0.99	0.98	130
Atopic dermatitis	0.98	0.99	0.99	130
Basal cell carcinoma (BCC)	0.98	0.99	0.99	125
Melanocytic nevi (NV)	0.98	0.97	0.98	110
Benign keratosis-like lesions	0.98	0.94	0.96	149
Psoriasis pictures lichen planus and related	1.00	1.00	1.00	134
Seborrheic keratosis and benign tumours	0.99	1.00	0.99	99
Tinea ringworm and other fungal infections	1.00	0.99	1.00	133

Next, for the Xception model, Figure 5 shows 50 epochs of training for training/validation and loss/accuracy. The base Xception model was unfrozen

after fine-tuning for an extra 20 epochs, with training and validation loss/accuracy after 70 epochs as shown in Figure 6. For the test set, the errors by class are shown in Figure 7. The evaluation metrics of the trained model show that loss = 0.19959 and accuracy = 0.99143. Table 2 shows the class-wise performance on the Xception model.

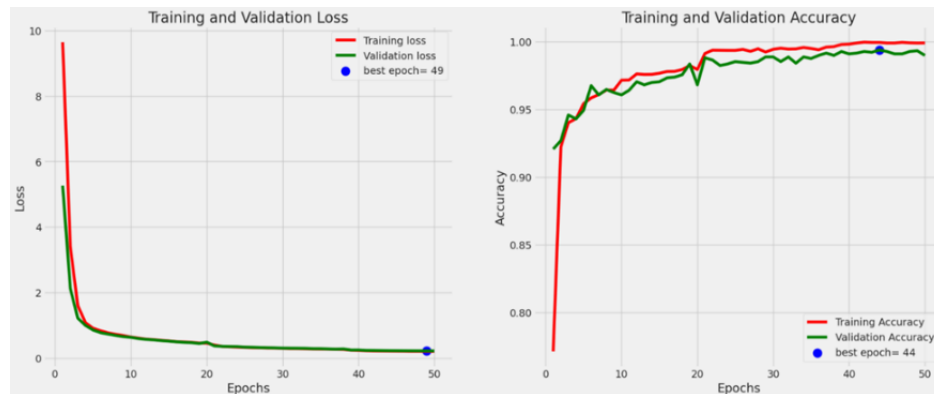


Figure 5 Training and validation loss and accuracy for 50 epochs on Xception.



Figure 6 Fine-tune training and validation loss and accuracy for 70 epochs on Xception.

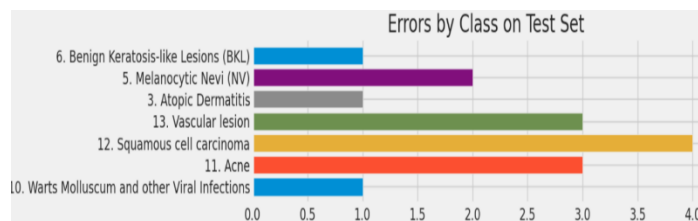
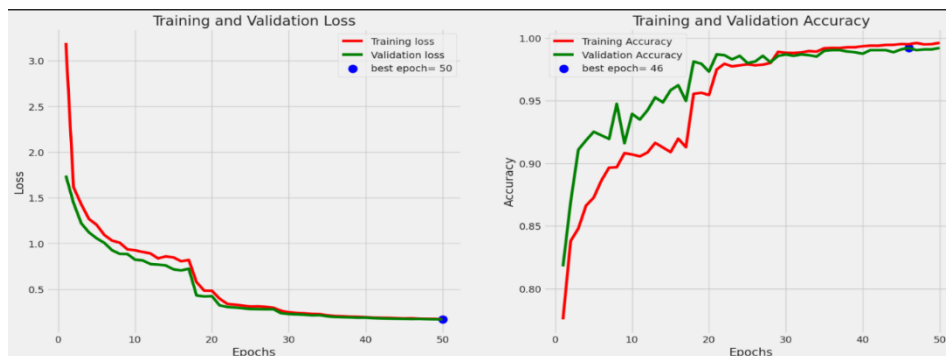


Figure 7 Errors by class set on Xception.

Table 2 Class-wise performance on Xception.

Classes	Precision	Recall	F1	Support
Eczema	1.00	1.00	1.00	129
Warts molluscum and other viral infections	1.00	0.99	1.00	114
Acne	1.00	0.98	0.99	131
Squamous cell carcinoma	1.00	0.97	0.98	121
Vascular lesion	0.99	0.97	0.98	112
Dermatofibroma	0.96	1.00	0.98	133
Melanoma	0.98	1.00	0.99	130
Atopic dermatitis	0.99	0.99	0.99	130
Basal cell carcinoma (BCC)	0.98	1.00	0.99	125
Melanocytic nevi (NV)	1.00	0.98	0.99	110
Benign keratosis-like lesions (BKL)	1.00	0.99	1.00	149
Psoriasis pictures lichen planus and related	1.00	1.00	1.00	134
Seborrheic keratosis and other benign tumours	1.00	1.00	0.99	99
Tinea ringworm and other fungal infections	0.99	1.00	0.99	133

In terms of DenseNet201 models, like other models, Figure 8 shows training for 50 epochs of training/validation, loss/accuracy. The base Xception model was unfrozen after fine-tuning for an extra 20 epochs showing training and validation loss/accuracy after 70 epochs, as shown in Figure 9. The DenseNet201 model demonstrated an impressive accuracy of 98.74%, highlighting its efficacy in precisely categorizing skin disease images. Figure 10 shows the class-wise errors on DenseNet-201. Table 3 shows the class-wise performance on the DenseNet-201 model.

**Figure 8** Training and validation loss and accuracy for 50 epochs on DenseNet-201.

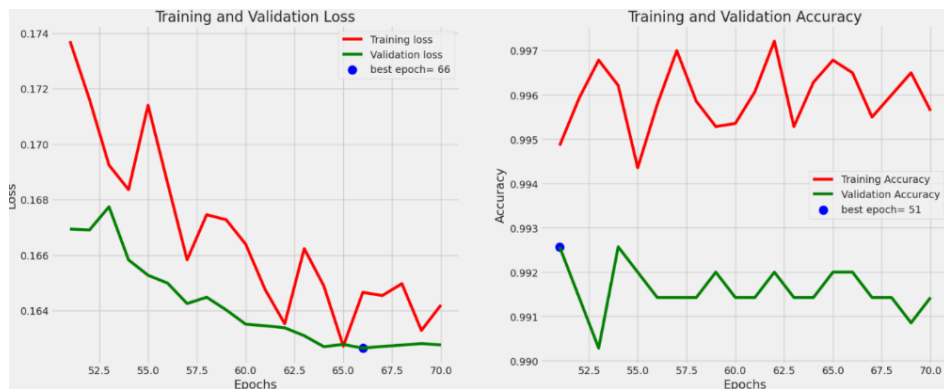


Figure 9 Fine-tune training and validation loss and accuracy for 70 epochs on DenseNet-201.

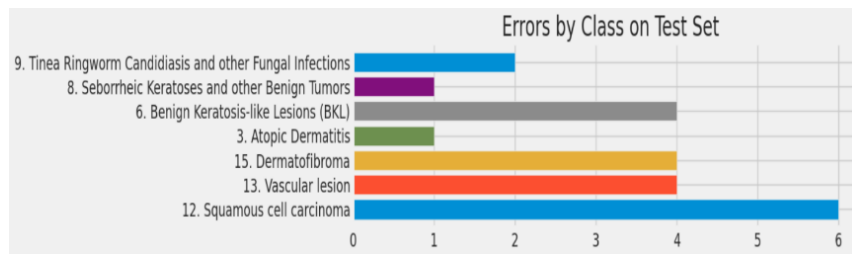


Figure 10 Errors by class set on DenseNet-201.

Table 3 Class-wise result of DenseNet-201.

Classes	Precision	Recall	F1	Support
Eczema	0.99	1.00	1.00	133
Warts molluscum and other viral infections	0.98	1.00	0.99	131
Acne	1.00	1.00	1.00	129
Squamous cell carcinoma	0.99	0.95	0.97	125
Vascular lesion	0.97	0.97	0.97	130
Dermatofibroma	1.00	0.97	0.99	149
Melanoma	1.00	1.00	1.00	110
Atopic dermatitis	1.00	0.99	1.00	133
Basal cell carcinoma (BCC)	0.98	1.00	0.99	112
Melanocytic nevi (NV)	0.95	1.00	0.98	99
Benign keratosis-like lesions (BKL)	0.95	0.97	0.96	121
Psoriasis pictures lichen planus and related	0.99	1.00	1.00	134
Seborrheic keratosis and other benign tumours	1.00	0.99	1.00	114
Tinea ringworm and other fungal infections	1.00	0.98	0.99	130

Our research emphasized furthering transfer learning through fine-tuning, following established studies, on a balanced data set with strict ethical considerations taken directly from the medical field. It was this careful attitude

that helped us to optimize the chosen CNN model and increase the accuracy considerably compared to earlier research.

As indicated in Table 4, the Xception model performed best, with test set accuracy as high as 99.14%. By comparison, DenseNet201 realized an accuracy of 98.74%, and InceptionV3 followed with an accuracy of 98.46%. These results demonstrate that our fine-tuning approach was very effective, especially in the detection of complex skin diseases. Moreover, the training loss values support the strength of our models, where Xception always remained at a lower loss of 0.19959, proving that it is efficient in learning from the data.

Table 5 compares our study with previous studies. This comparison shows that our model did not only perform more accurately than previous studies but also did so using a much larger and more diversified data set. For instance, while Vgg16 combined with Xgboost realized an accuracy of 99.10% on a far smaller data set of 1,416 images, our Xception model was able to surpass this with 99.14% on a data set with 17,500 images in 14 classes.

This study makes significant contribution to the field in the form of a more correct and complete solution for skin disease detection, outperforming the available methodologies. The increased data set and number of classes also give importance and strength to our model in helping medical professionals diagnose a wide variety of skin conditions.

Table 4 Proposed models' accuracy on test set with training loss and accuracy value.

Models	Accuracy on test set (%)	Loss value after training	Accuracy value after training
Xception	99.14%	0.19959	0.99143
DenseNet201	98.74%	0.16492	0.98743
InceptionV3	98.46%	0.25455	0.98457

Table 5 Comparison with other researcher work.

Reference study	Data set	No. of classes	Best method	Accuracy (%)
[10]	1,416	2	Vgg16 + Xgboost	99.10%
[16]	12,900	12	ResNet50	83.62%
[17]	1,800	2	DenseNet121	84.64%
This study*	17,500	14	Xception	99.14%

5 Conclusion

This study highlighted the application of deep feature extraction and transfer learning for the identification of skin diseases from a diverse global skin disease image collection. Three widely recognized and updated deep CNN architectures – DenseNet201, Xception, and Inception V3 – were employed for deep feature

extraction and transfer learning. The data set used in our experimental work stands out due to its substantial quantity of example images and a diverse range of classes compared to other reviewed studies. Xception demonstrated exceptional performance, achieving a remarkable 99.14% accuracy rate, outperforming other algorithms considered in our analysis of existing research articles on skin disease detection. In future work we will aim to further enhance classification accuracy by incorporating various CNN models. This research contributes to the overarching goal of developing an automated system capable of real-time skin disease diagnosis through the scanning of personal skin images. Such an automated system has the potential to assist the medical field in promptly identifying and addressing skin issues. In the future, we will use our approach to create an automated disease detection app with precision medicine.

References

- [1] Hosny, K.M., Kassem, M.A. & Foad, M.M., *Classification of Skin Lesions using Transfer Learning and Augmentation with Alex-Net*, PloS One, **14**(5), e0217293, May, 2019.
- [2] Bitto, A.K. & Mahmud, I., *Multi Categorical of Common Eye Disease Detect using Convolutional Neural Network: A Transfer Learning Approach*, Bulletin of Electrical Engineering and Informatics, **11**(4), pp.2378-2387, 2022.
- [3] Hosny, K.M., Kassem, M.A. & Foad, M.M., *Skin Melanoma Classification using ROI and Data Augmentation with Deep Convolutional Neural Networks*, Multimedia Tools and Applications, **79**, pp.24029-24055, June, 2020.
- [4] El Gannour, O., Hamida, S., Lamalem, Y., Cherradi, B., Saleh, S. & Raihani, A., *Enhancing Skin Diseases Classification Through Dual Ensemble Learning and Pre-trained CNNs*, International Journal of Advanced Computer Science and Applications, **14**(6), pp.436-445, 2023.
- [5] Jahanifar, M., Tajeddin, N.Z., Koohbanani, N.A., Gooya, A. & Rajpoot, N., *Segmentation of Skin Lesions and Their Attributes using Multi-scale Convolutional Neural Networks and Domain Specific Augmentations*, arXiv preprint, 2018.
- [6] Bakkouri, I. & Afdel, K., *Computer-Aided Diagnosis (CAD) System based on Multi-Layer Feature Fusion Network for Skin Lesion Recognition in Dermoscopy Images*, Multimedia Tools and Applications, **79**(29-30), pp.20483-20518, 2020.
- [7] Abunadi, I. & Senan, E.M., *Deep Learning and Machine Learning Techniques of Diagnosis Dermoscopy Images for Early Detection of Skin Diseases*, Electronics, **10**(24), 3158., Dec. 2021.
- [8] Shaheen, H. & Singh, M.P., *Multiclass Skin Cancer Classification using Particle Swarm Optimization and Convolutional Neural Network with*

- Information Security, Journal of Electronic Imaging*, **32**(4), pp.042102-042102, 2023.
- [9] Raghavendra, P.V., Charitha, C., Begum, K.G. & Prasath, V.B.S., *Deep Learning-based Skin Lesion Multi-class Classification with Global Average Pooling Improvement*, *Journal of Digital Imaging*, **36**(5), pp.2227-2248, 2023.
 - [10] Thanka, M.R., Edwin, E.B., Ebenezer, V., Sagayam, K.M., Reddy, B.J., Günerhan, H. & Emadifar, H., *A Hybrid Approach for Melanoma Classification using Ensemble Machine Learning Techniques with Deep Transfer Learning*, *Computer Methods and Programs in Biomedicine*, **3**, 100103, 2023.
 - [11] Alani, A.A., Altameemi, H.G., Azeez Asmael, A.A. & Al-Obaidi, M.A., *An Effective Multiclass Human Skin Lesion Diagnosis System based on Convolutional Neural Networks*, *Automatic Control and Computer Sciences*, **57**(2), pp.135-142, May, 2023.
 - [12] Adegun, A.A. & Viriri, S., *FCN-based Densenet Framework for Automated Detection and Classification of Skin Lesions in Dermoscopy Images*, *IEEE Access*, **8**, pp.150377-150396, 2020.
 - [13] Mehmood, A., Gulzar, Y., Ilyas, Q.M., Jabbari, A., Ahmad, M. & Iqbal, S., *SBXception: A Shallower and Broader Xception Architecture for Efficient Classification of Skin Lesions*, *Cancers*, **15**(14), 3604., July, 2023.
 - [14] Goyal, M., Yap, M.H. & Hassanpour, S., *Multi-class Semantic Segmentation of Skin Lesions via Fully Convolutional Networks*, arXiv preprint, 2017.
 - [15] Balaji, V.R., Suganthi, S.T., Rajadevi, R., Kumar, V.K., Balaji, B.S. & Pandiyan, S., *Skin Disease Detection and Segmentation using Dynamic Graph Cut Algorithm und Classification Through Naive Bayes Classifier*, *Measurement*, **163**, 107922, 2020.
 - [16] Al-Masni, M.A., Kim, D.H. & Kim, T.S., *Multiple Skin Lesions Diagnostics via Integrated Deep Convolutional Networks for Segmentation and Classification*, *Computer methods and programs in biomedicine*, **190**, 105351, July, 2020.
 - [17] Sonawane, R.S., *Skin-cancer Classification using Deep Learning with DenseNet and VGG with Streamlit-framework Implementation*, PhD dissertation, National College of Ireland, Dublin, 2023.
 - [18] Usama, M., Naeem, M.A. & Mirza, F., *Multi-class Skin Lesions Classification using Deep Features*, *Sensors*, **22**(21), 8311, 2022.
 - [19] Gairola, A.K., Kumar, V., Sahoo, A.K., Diwakar, M., Singh, P. & Garg, D., *Multi-feature Fusion Deep Network for Skin Disease Diagnosis*, *Multimedia Tools and Applications*, 1-26, 2024.
 - [20] Purni, J.T. & Vedhapriyavadhana, R., *EOSA-Net: A Deep Learning Framework for Enhanced Multi-class Skin Cancer Classification using*

- Optimized Convolutional Neural Networks*. Journal of King Saud University-Computer and Information Sciences, **36**(3), 102007, 2024.
- [21] *Skin Diseases Image Dataset*, (2021, August 16). Kaggle. <https://www.kaggle.com/datasets/ismailpromus/skin-diseases-image-dataset>
 - [22] Tschandl, P., *The HAM10000 Dataset, a Large Collection of Multi-Source Dermatoscopic Images of Common Pigmented Skin Lesions [Dataset]*, In Harvard Dataverse, 2018. <https://doi.org/10.7910/dvn/dbw86t>
 - [23] Center for Artificial Intelligence in Medicine & Imaging, *Shared datasets*. <https://aimi.stanford.edu/shared-datasets> (11 September 2024)
 - [24] Niloy, G.M., Bitto, A.K., Biplob, K.B.M.B., Sammak, M.H., Das, A. & Hridoy, G.G., *MobileNet-Eye: An Efficient Transfer Learning for Eye Disease Classification*. In 2024 International Conference on Advances in Computing, Communication, Electrical, and Smart Systems (iCACCESS) (pp. 01-05), IEEE, March, 2024.
 - [25] Bitto, A.K., Bijoy, M.H.I., Yesmin, S., Mahmud, I., Mia, M.J., Biplob, B. & Md, K.B., *Tumor-Net: Convolutional Neural Network Modeling for Classifying Brain Tumors from MRI Images*, International Journal of Advances in Intelligent Informatics, **9**(2), pp.148-160, 2023.
 - [26] Bitto, A.K., Bijoy, M.H.I., Yesmin, S. & Mia, M.J., *Tumor-TL: A Transfer Learning Approach for Classifying Brain Tumors from MRI Images*, In International Conference on Machine Intelligence and Emerging Technologies, pp. 177-186, 2023.