ICIC Express Letters
Volume **18**, Number **6**, June **2024**
 ISSN 1881-803X
pp. 563–573

ADDRESSING IMBALANCE ISIC-2019 DATASET IN DERMOSCOPIC PIGMENTED SKIN LESION CLASSIFICATION

Erwin Setyo Nugroho[1,2], Igi Ardiyanto[1] and Hanung Adi Nugroho[1,*]

[1]Department of Electrical and Information Engineering
Engineering Faculty
Universitas Gadjah Mada
Kompleks Grafika, Yogyakarta 55281, Indonesia
igi@ugm.ac.id; *Corresponding author: adinugroho@ugm.ac.id

[2]Department of Informatics
Politeknik Caltex Riau
Jl. Umbansari No. 1 Rumbai, Pekanbaru, Riau 28265, Indonesia
erwinsn@pcr.ac.id

Received June 2023; accepted September 2023

Abstract. *Skin cancer, particularly melanoma, is a life-threatening condition that requires early detection and prompt treatment to reduce mortality rates. Dermoscopic images offer a non-invasive and cost-effective method for examining pigmented skin lesions. However, accurate analysis is challenging due to the lack of standardized colours, image capture settings, and artefacts. Deep learning models, such as Convolutional Neural Networks (CNNs), have shown promising Computer-Aided Diagnosis (CAD) results by automatically extracting features from medical images. Nonetheless, the performance of these models heavily relies on the quality and balance of the training dataset. This study addresses the imbalance issue within the ISIC-2019 dataset, which contains dermoscopic images of pigmented skin lesions. Three pre-trained CNNs (Inception-v3, Xception, and DenseNet-201) were chosen to implement the augmentation scenario. The dataset underwent several preprocessing steps, including duplicate detection, data cleaning, and resizing. Additionally, data augmentation techniques were applied to balancing the distribution of images across different classes. Experimental results demonstrated the effectiveness of the proposed method in improving the classification performance of the pre-trained CNNs. These findings underscore the significance of dataset preparation and augmentation techniques in overcoming challenges posed by imbalanced datasets. The results validate the effectiveness of data preprocessing and augmentation techniques in achieving higher classification accuracy.*
Keywords: Imbalanced dataset, Skin lesion, ISIC-2019, Augmentation

1. **Introduction.** Pigmented skin lesions, characterized by abnormal pigmentation compared to the surrounding normal skin, encompass a range of conditions. Among these, melanoma stands out as the most lethal variant of skin cancer affecting pigmented skin lesions [1]. While constituting only around 1% of skin cancer cases, melanoma is responsible for a disproportionate number of fatalities within the realm of skin cancer. The peril of melanoma arises from its potential to metastasize to distant organs if not promptly diagnosed and treated. Early intervention in melanoma cases is pivotal in mitigating mortality rates. Fortunately, advancements in treatment modalities have led to a notable reduction in melanoma-related deaths over the last decade (2011 to 2020). Rates of decline have been approximately 5% per year for adults under 50 and about 3% annually for individuals aged 50 and above [2].

However, despite these encouraging trends, challenges persist. Over recent decades, the incidence of melanoma has exhibited significant fluctuations across various age groups.

DOI: 10.24507/icicel.18.06.563

Notably, a continuous annual rise of approximately 1% has been observed in women aged 50 and above from 2015 to 2019, while rates for men in the same age bracket have stabilized. This divergence underscores the intricate nature of melanoma's dynamics within different demographic segments.

The motivation behind this study stems from the need to address a critical impediment in the domain of pigmented skin lesion classification – the issue of imbalanced datasets. Dermatologists rely extensively on dermoscopic images, offering a non-invasive, cost-effective means to assess skin tissue [3]. However, the efficacy of automated image analysis techniques, particularly those employing Convolutional Neural Networks (CNNs), is hampered by the lack of standardized colors, methodologies, and image capture settings, leading to variations in outcomes. Moreover, the presence of numerous artifacts such as blood vessels, skin hair, dark spots, and even extraneous marks further complicates the accurate identification and categorization of lesions. These challenges are amplified by the absence of direct physician examination, making it difficult for researchers to explore diverse techniques that could yield optimal detection outcomes.

Deep learning, particularly the use of CNNs, has experienced remarkable progress in Computer-Aided Diagnostic (CAD) systems. CNNs excel in extracting salient features from medical images, such as radiographs and CT scans, facilitating the detection of abnormalities and signs of diseases with remarkable accuracy, sometimes surpassing human diagnostic capabilities [4,5]. This advancement has considerably enhanced the precision and efficiency of medical diagnoses, presenting an invaluable tool for medical practitioners to bolster clinical decision-making.

The efficacy of employing CNNs for classifying pigmented skin lesions, however, hinges on the availability of a comprehensive dataset. The occurrences of various diseases within the realm of pigmented skin lesions are unevenly distributed, leading to dataset imbalances that pose a significant challenge for deep learning-based classification [6]. Many augmentation techniques are currently being developed, including resampling techniques [7], algorithmic techniques [8], and data level techniques [9]. The choice of technique depends on the specifics of the data set and the problem. In this research, the resampling technique is used due to the simplicity of the process. The resampling technique in this research targets the amount of data in the largest class, so that data in small classes is oversampled.

In this context, this research endeavor contributes significantly to the field of dermatology and skin cancer diagnosis. It outlines a robust methodology aimed at addressing the complexities of imbalanced datasets and enhancing the performance of pre-trained CNNs in accurately classifying pigmented skin lesions. The subsequent sections of this paper elaborate on the materials and methods employed (Section 2), present the outcomes of applying the proposed model to five pre-trained CNNs (Section 3), and conclude by summarizing the critical insights and implications of this research.

2. **Materials and Methods.** This section will explain the dataset, pre-trained CNN, augmentation scenario, and experimental setup used in this research and its evaluation.

2.1. **Pre-trained CNN models.** A pre-trained CNN model is a CNN model trained on a large dataset before being used. It learns to extract features from images and can recognize various visual patterns. Pre-trained models save time and resources compared to training from scratch. They achieve high accuracy by leveraging large-scale datasets like ImageNet. Researchers and developers can use pre-trained models for tasks like image classification, object detection, and feature extraction. Fine-tuning allows adapting the model to new datasets. Overall, pre-trained CNN models are valuable resources for efficiently solving image-related problems. In this study, all pre-trained CNN models have been trained using ImageNet.

We have selected three frequently utilized pre-trained CNNs. The selection of the specific model is not critical in this case, as we aim to assess the effects of augmentation rather than striving for peak performance. Considering resource limitations, we opted for models with hyper-parameter counts ranging from 20 to 25 million. The selected models, including Inception-v3, Xception, and DenseNet-201, are presented in Table 1.

Table 1. A description of the employed architecture

Model	Architect	Depth	Parameters	Input size
Inception-v3	Szegedy et al. [10]	189	23.9 M	299 × 299
DenseNet-201	Huang et al. [11]	402	20.2 M	299 × 299
Xception	Chollet [12]	81	22.9 M	299 × 299

Inception-v3 is a powerful model with multiple pathways for capturing features at different scales. Deep architecture and efficient design allow for effective learning of hierarchical representations. Inception-v3's design revolves around the inception module, which combines convolutional layers of different sizes to recognize patterns and objects. The deep architecture of stacked inception modules enables the model to capture low-level and high-level features. Overall, Inception-v3 excels in image-related tasks and is widely used in computer vision applications.

DenseNet-201 possesses dense connection, a deep architecture, feature concatenation, effective parameter use, pre-training on massive datasets, and adaptability. DenseNet-201 effectively uses parameters, resulting in a compact model with enhanced performance. Pre-training on massive datasets improves precision and permits transfer learning. It is flexible and adaptable to some computer vision jobs. Overall, DenseNet-201 provides excellent network performance and efficient information flow.

Essential characteristics of the Xception model include depthwise separable convolutions for efficient feature capture, extreme depth for hierarchical representations, linear bottleneck structure for parameter reduction, pre-training on large-scale datasets for generalization, versatility for various computer vision tasks, and high accuracy in image-related tasks. Xception is a practical framework for image recognition and computer vision applications.

2.2. **ISIC dataset.** The International Skin Imaging Collaboration (ISIC) is a multinational organization aiming to improve the early detection and diagnosis of melanoma and other skin cancers using digital imaging and Artificial Intelligence (AI) technology. It was formed in 2016 and has since become an industry leader in dermatology and skin imaging. Since 2016, ISIC has made available the dermoscopic dataset, updated annually for their challenges. Table 2 and Table 3 detail the datasets that ISIC has made available.

In terms of quantity, there is an increase in data every year, but the number of disease classes varies. The ISIC-2020 dataset has many data and disease classes, but the number

Table 2. The number of images contained in the ISIC 2016-2020 databases

Databases	Training	Testing	Class
ISIC-2016	900	379	2
ISIC-2017	2,000	600	3
ISIC-2018	10,015	1,512	7
ISIC-2019	25,331	8,238	8
ISIC-2020	33,126	10,982	2

TABLE 3. ISIC 2016-2020 class distribution training data datasets

Class disease	2016	2017	2018	2019	2020
melanoma	173	374	1,113	4,522	584
melanocytic nevus	0	0	6,705	12,875	0
basal cell carcinoma	0	0	514	3,323	0
actinic keratosis	0	0	327	867	0
benign keratosis	0	0	1,099	2,624	0
dermatofibroma	0	0	115	239	0
vascular lesion	0	0	142	253	0
sqs cell carcinoma	0	0	0	628	0
benign	727	1,372	0	0	0
seborrheic keratosis	0	254	0	0	135
atypical melanocytic	0	0	0	0	1
cafe-au-lait macule	0	0	0	0	1
lentigo nos	0	0	0	0	44
lichenoid keratosis	0	0	0	0	37
nevus	0	0	0	0	5,193
solar lentigo	0	0	0	0	7
other/unknown	0	0	0	0	27,124
Total	900	2,000	10,015	25,331	33,126

of unknown images accounts for over 80% of the total data, although they can be categorized as benign. The ISIC-2020 dataset is more commonly used for binary classification between malignant melanoma and benign cases. Despite being imbalanced, the ISIC-2019 dataset has the largest number of classes and sufficient data for each class. Considering the distribution of the above datasets, this study uses the ISIC-2019 dataset.

The ISIC-2019 dataset is freely available to the public for non-commercial use under the CC-BY-NC license. The dataset includes two CSV files: one containing metadata, which provides general information about patients and their lesions, such as age, gender, and anatomical location, and the other containing the confirmed diagnosis of each lesion, serving as the ground truth for analysis.

2.3. **Dataset preprocessing and augmentation scenario.** In this section, we detail our dataset preprocessing and augmentation strategy, geared towards cultivating a robust dataset that optimally supports our model's effectiveness and understanding of skin lesions. This multi-step pipeline is designed to fortify our model.

1) Duplicate Detection for Uniqueness: We initiate with a stringent duplicate detection process to maintain dataset integrity. Scrutinizing and comparing each image, we eliminate duplicates, ensuring distinct data for training and testing, bolstering evaluation accuracy.
2) Data Cleaning for Metadata Consistency: Ensuring precision, we meticulously inspect image metadata. Images lacking lesion IDs are removed, ensuring alignment with our focus. This diligence guarantees faithful representation of skin lesion categories.
3) Resizing for Model Compatibility: Adapting images to pre-trained CNN model requisites, we standardize dimensions. This promotes efficient feature extraction and convergence, pivotal for accurate classification.
4) Augmentation Strategy: Our augmentation framework, integral to our approach, diversifies training data. Leveraging an Image Data Generator, diverse transformations enable effective model generalization across scenarios. Transformations include up to 180-degree rotation, 0.1 shifting in height and width, 0.1 zooming, horizontal/vertical

flipping, brightness adjustment (0.9 to 1.1 range), and "nearest" fill mode. Real-world variations are introduced, aiding adaptability.

5) Balanced Class Exposure and Evaluation: Employing a batch size of 20, we manage augmented images efficiently, targeting a final count of 9200, matching the largest class (NV). Reserving 800 images (100 per class) for testing, our splitting strategy and augmentation address class distribution bias, ensuring equitable learning. Summary results are presented in Table 4.

Table 4. Data splitting and augmentation scenarios

Class	**Initial data**	**No ID lesion**	**Testing dataset**	**Training dataset**	**Augmented training dataset**
NV	12,875	3,647	100	9,128	9,128
MEL	4,522	495	100	3,927	9,204
BCC	3,323	138	100	3,085	9,220
BKL	2,624	436	100	2,088	9,202
AK	867	36	100	731	9,060
SCC	628	24	100	504	8,606
DF	239	11	100	128	8,436
VASC	253	39	100	114	7,918
Total	**25,331**	**4,826**	**800**	**19,705**	**70,774**

Note that NV: melanocytic nevus, MEL: melanoma, BCC: basal cell carcinoma, BKL: benign keratosis, AK: actinic keratosis, SCC: squamous cell carcinoma, DF: dermatofibroma, VASC: vascular lesion.

To gauge augmentation's impact, we plan comprehensive experiments, contrasting model performance with and without augmentation. Quantifying enhancements in accuracy, convergence, and generalization, this assessment validates our approach's efficacy.

2.4. **Experiment setup.** This research investigated various pre-trained CNN architectures for classifying pigmented skin lesions, as outlined in Table 1. All models were configured with the same set of default hyperparameters by Keras. ADAM optimization technique was used with a learning rate of 0.01, a dropout rate of 0.25, 50 epochs, and a batch size of 20.

The experimental procedure follows the diagram in Figure 1. A dataset comprising 25,331 images of pigmented skin lesions across eight disease classes is utilized as input. A preprocessing stage is applied to the data before feeding the images into the pre-trained CNN. This stage includes duplicate removal and resizing of the images. Since the images in the ISIC-2019 dataset have varying sizes, they are resized to match the input size of the model. Data augmentation techniques are also employed to increase the number of input samples during the training process.

This step also balances the number of images in each class. Finally, the data is split into training and testing sets. Out of the original 800 data points, 800 are set aside for testing purposes, while the remaining data is used for training with 5-fold cross-validation. Following the preprocessing stage, the training phase commences using ImageNet transfer learning for each CNN architecture. Once satisfactory learning results are achieved, testing is performed using the previously separated 800 data points. Using ImageNet transfer learning accelerates the training process and improves the performance of the models.

The training procedure took place using JupiterHub, a platform that facilitates the design and implementation of CNN architectures. The framework used for training was deployed on the JupiterHub platform. Due to the computationally demanding nature

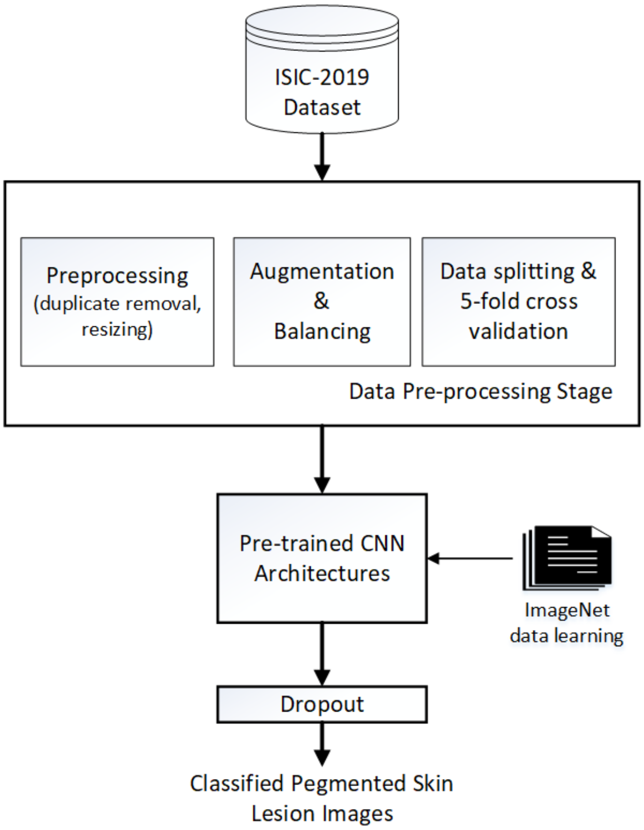

FIGURE 1. Schematic of the experiment

of CNN training, all experiments were performed on workstations equipped with the following specifications: an Intel (R) Core i9-10900K 3.7 GHz processor, 128 GB of RAM, an NVIDIA RTX 3080 11 GB GPU, and the Linux Ubuntu 18.04 operating system.

2.5. **Evaluation.** The quality of a learning algorithm is determined by analyzing its performance on test data. The evaluation matrix design starts with the confusion matrix, as presented in Figure 2. Commonly used performance evaluation metrics for classification include sensitivity (SEN), specificity (SPE), accuracy (ACC), precision (PREC), and the Area under the Receiver Operating Characteristic (ROC) curve (AUC) [13].

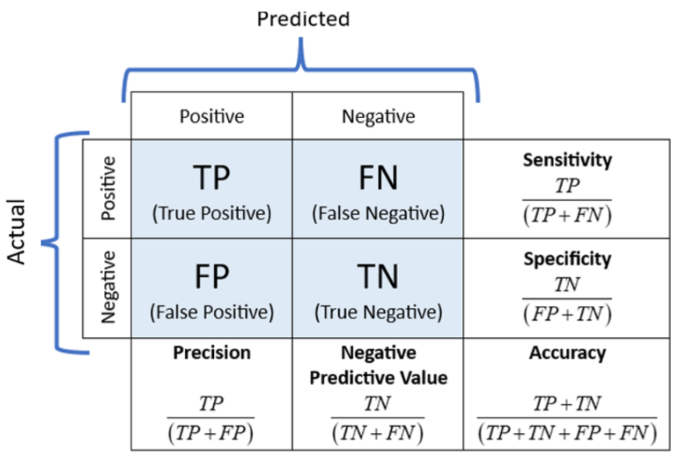

FIGURE 2. Confusion matrix with actual vs prediction

3. **Result and Discussion.** This section presents and discusses the outcomes of applying the experimental scheme to chosen pre-trained CNNs using both the original and augmented datasets. Additionally, a comparison is made with other existing research findings. The limitations and potential directions for future research are addressed towards the end.

3.1. **Result in training and testing models.** After training a selected CNN model with the ImageNet dataset to establish initial weights, the model is further trained using the ISIC-2019 dataset. The training process involves utilizing both the original dataset and the augmented dataset. The training accuracy and validation accuracy results of the Inception-v3, DenseNet-201, and Xception models with both the original and augmented datasets are presented in Figure 3. To avoid bias in the splitting of training and validation data, the model is trained using 5-fold cross-validation. The training process consists of 50 epochs.

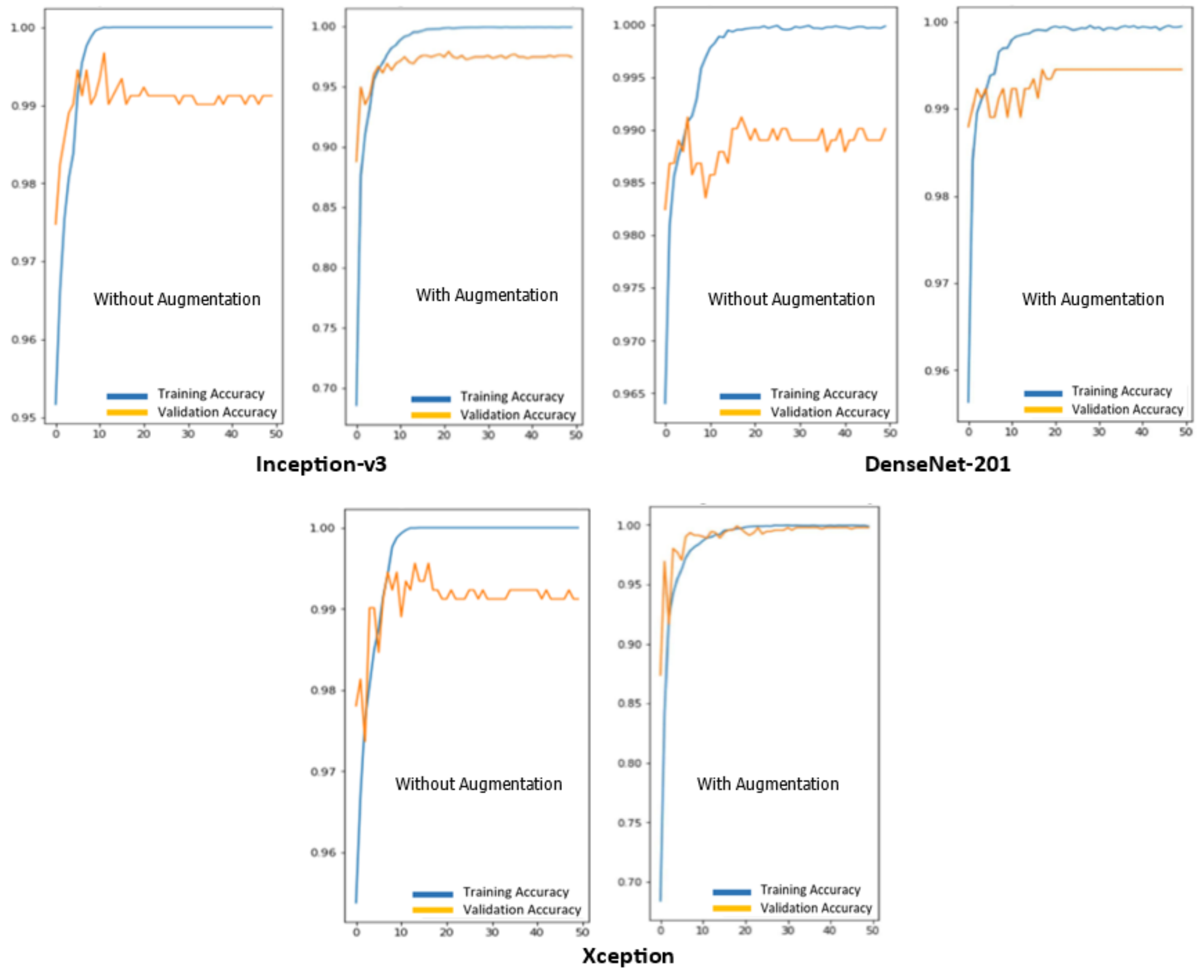

FIGURE 3. Training and validation accuracy of pre-trained CNN with and without augmentation. The X-axis is the number of epochs, and the Y-axis is the accuracy value.

A model's ability to accurately identify new objects is greatly influenced by its ability to fit well. CNN models are considered a good fit when the validation accuracy closely follows the training accuracy. Figure 3 shows that all models fit well with the original and augmented datasets. However, it is noticeable that the models trained with augmented data have validation accuracy results that are more similar to the training results than those trained with the original dataset. This outcome demonstrates the positive impact of augmentation on the models. Furthermore, based on statistical evaluation of the results of the callback function in Tensorflow, the model reaches its optimal and stable state starting from the 20th epoch. The state is confirmed in the graph shown in Figure 3.

In Figure 4, we are presented with a confusion matrix that compares the performance of a pre-trained Convolutional Neural Network (CNN) with and without augmentation. The confusion matrix provides insights into the classification results by displaying the predicted labels against the actual labels for each class. The confusion matrix is divided into "Without Augmentation" and "With Augmentation". Each section represents the performance of the CNN model under different conditions. The rows in the matrix represent the actual labels, while the columns represent the predicted labels.

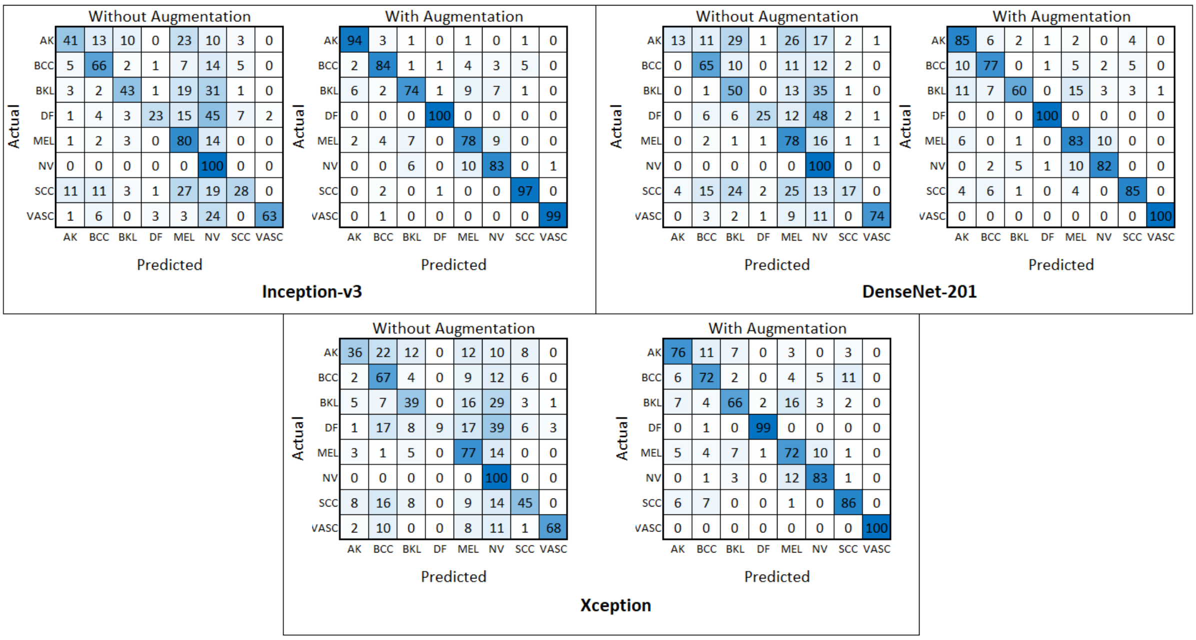

FIGURE 4. Confusion matrix testing of pre-trained CNN with and without augmentation

Analyzing the confusion matrix of the model without augmentation reveals that the Melanocytic Nevus (NV) class achieves 100% accuracy across all three models. This condition can be attributed to the dominance of the NV class within the original dataset, accounting for 12,875 images or 50.8% of the overall data. As a result, the NV class is perfectly detected by all three models. However, this imbalance adversely affects other classes, especially those with limited images. On the other hand, the confusion matrix of the model with augmentation demonstrates accuracy distributed across all classes. This condition indicates an improvement in model performance through augmentation, which helps balance the representation of images between classes.

By comparing the results of the two sections, we can assess the impact of augmentation on the model's performance. The "With Augmentation" section generally shows better results regarding increased true positives and reduced misclassifications. However, it is essential to note that the specific augmentation techniques used are not provided in the given information. Augmentation may entail picture modifications such as rotation, scaling, inversion, or the addition of noise. The success of augmentation is dependent on the particular dataset and the approaches chosen.

Figure 4 depicts, in conclusion, the effect of augmentation on the performance of a pre-trained CNN. The results indicate that augmentation enhances the model's capacity to categorize the various skin lesion types, as seen by a rise in true positives and a decrease in false positives. Table 5 shows the total performance of the examination.

Table 5. Performance of pre-trained CNN with and without augmentation

Model	SEN (%)	SPE (%)	PRE (%)	ACC (%)	F1 (%)	AUC
Without augmentation						
Inception-v3	55.50	90.40	65.06	55.50	54.10	0.50
DenseNet-201	52.75	89.41	64.06	52.75	49.29	0.50
Xception	55.13	64.97	64.71	55.13	52.22	0.50
With augmentation						
Inception-v3	88.63	98.23	88.53	88.63	88.53	0.94
DenseNet-201	84.00	97.40	84.60	84.00	83.86	0.90
Xception	81.75	96.98	81.78	81.75	81.69	0.89

Table 5 compares the performance of CNN models pre-trained with and without data augmentation. The outcomes indicate the value of augmentation strategies for enhancing the performance of models. Without augmentation, the models' values for various parameters, including sensitivity, specificity, precision, accuracy, F1 score, and AUC, were considerably lower. However, when augmentation was done, all models demonstrated significant performance enhancements. The enhanced models exhibited considerably greater sensitivity, specificity, precision, accuracy, F1 score, and Area under the Curve (AUC) than their non-augmented counterparts. This accomplishment demonstrates that augmentation approaches improved the models' capacity to generalize and generate more accurate predictions on unobserved data.

These findings highlight the importance of data augmentation as a valuable tool in training CNN models. Augmentation helps to diversify the training data by introducing variations, leading to better model performance and increased robustness. In practical applications of CNN models, where accuracy and performance are crucial, employing data augmentation techniques can be a beneficial strategy to enhance the model's overall performance and achieve more reliable results.

3.2. **Comparison with existing research.** This section will compare the results of this study with those of previous studies. The comparison is restricted to classification studies utilizing the ISIC-2019 eight-class dataset with an augmentation process, as Table 6 shows. In terms of sensitivity, specificity, precision, accuracy, and F1 score, the models in the current study consistently beat the other methods/models when comparing performance measures. The Inception-v3 model earned the most excellent Area under the Curve (AUC) score of 0.94.

Table 6. Comparison of performance with existing research

Author	Method/model	Performance metrix					
		SEN (%)	SPE (%)	PRE (%)	ACC (%)	F1 score (%)	AUC
Molina-Molina et al. [14]	DenseNet-201	66.45	97.85	**91.61**	**97.35**	n/a	n/a
Kassem et al. [15]	GoogleNet	79.80	97.00	80.36	94.92	80.07	n/a
Liu et al. [16]	Clinical-Inspired	53.80	97.40	n/a	64.00	n/a	0.91
Cauvery et al. [17]	Ensemble CNN	62.00	98.00	73.00	81.00	56.00	n/a
Ours	Inception-v3	**88.63**	**98.23**	88.53	88.63	**88.53**	**0.94**
Ours	DenseNet-201	84.00	97.40	84.60	84.00	83.86	0.90
Ours	Xception	81.75	96.98	81.78	81.75	81.69	0.89

Molina-Molina et al. [14] address the issue of imbalanced classes by employing three classifiers with a linear plurality vote. Although this approach yields high accuracy and precision, it falls short in improving sensitivity, a crucial parameter in the medical field that cannot be overlooked. Furthermore, this method necessitates extensive computational effort. While Liu et al. [16]'s augmentation process lacks specific explanations regarding the methods and procedures employed, the obtained performance results were also unsatisfactory. Cauvery et al. [17] tackle the problem of imbalanced classes by utilizing an online-augmentation policy. Although this method offers the advantage of not directly increasing the number of training images, it presents numerous disadvantages, including dependency on online connection, higher computational costs, reliance on input data quality, and the risk of overfitting. Like Kassem et al. [15]'s research, our study incorporates augmentation concepts where the number of images in each class is augmented to approach the number of images in the largest class. However, our research demonstrates several advantages, particularly regarding data cleaning and splitting processes. Notably, the testing data is guaranteed to remain separate from the augmented training data. In Table 6, our model outperforms Kassem et al.'s model.

These results highlight the effectiveness of the models presented in the current study. The utilization of data augmentation techniques further bolstered the models' performance. The findings suggest that these models have the potential to serve as valuable tools for accurate classification and discrimination of the relevant data.

Overall, the research conducted in the current study contributes significantly to the field by presenting robust models that outperform existing approaches. The models' demonstrated superiority in performance metrics underscores their potential impact and usefulness in practical applications within the given domain.

4. Conclusions.

In summary, this study has demonstrated the efficacy of training CNN models using the ImageNet dataset and subsequently fine-tuning them with the ISIC-2019 dataset. Augmentation played a pivotal role in achieving models that exhibited improved validation accuracy, aligning more closely with training accuracy. The application of augmentation not only boosted model performance but also contributed to a more balanced representation of images across different classes, as evident from the confusion matrix analysis. The conclusive evidence of augmentation's positive impact highlights its significance in optimizing CNN model performance for real-world applications.

The findings from this study underscore the importance of data augmentation in bolstering CNN model capabilities and attaining dependable outcomes in practical scenarios. By presenting robust models that surpass existing approaches, this research makes a valuable contribution to the field. The superior performance metrics achieved by these models underscore their potential utility and effectiveness within the designated domain.

This study identifies limitations and future research directions. The approach of using a single augmented image quantity might overlook optimal levels, while hyperparameter tuning could enhance model performance. Basic preprocessing techniques could be extended for better input quality. Incorporating specialized layers like attention mechanisms and exploring multi-modal approaches offer potential performance gains, albeit with increased training time. Addressing these aspects can advance the field and lead to more effective CNN models.

Acknowledgment. This research was sponsored by LPDP (Indonesia Endowment Fund for Education).

REFERENCES

[1] A. A. Adegun and S. Viriri, Deep learning-based system for automatic melanoma detection, *IEEE Access*, vol.8, pp.7160-7172, 2020.

[2] The American Cancer Society, *Key Statistics for Melanoma Skin Cancer*, 2022, https://www.cancer.org/cancer/types/melanoma-skin-cancer/about/key-statistics.html, Accessed on May 17, 2023.
[3] S. S. Venugopal, H. P. Soyer and S. W. Menzies, Results of a nationwide dermoscopy survey investigating the prevalence, advantages and disadvantages of dermoscopy use among Australian dermatologists, *Australas. J. Dermatol.*, vol.52, no.1, pp.14-18, 2011.
[4] A. Esteva et al., Dermatologist-level classification of skin cancer with deep neural networks, *Nature*, vol.542, no.7639, pp.115-118, 2017.
[5] T. J. Brinker et al., Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task, *Eur. J. Cancer*, vol.113, pp.47-54, 2019.
[6] P. Tschandl et al., Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: An open, web-based, international, diagnostic study, *Lancet Oncol.*, vol.20, no.7, pp.938-947, 2019.
[7] D. Elreedy and A. F. Atiya, A comprehensive analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance, *Inf. Sci.*, vol.505, pp.32-64, 2019.
[8] C. Elkan, The foundations of cost-sensitive learning, *Proc. of the 17th International Joint Conference on Artificial Intelligence (IJCAI'01)*, 2001.
[9] G. Weiss, K. McCarthy and B. Zabar, Cost-sensitive learning vs. sampling: Which is best for handling unbalanced classes with unequal error costs?, *International Conference on Data Mining*, pp.35-41, 2007.
[10] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, Rethinking the inception architecture for computer vision, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.2818-2826, 2016.
[11] G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, Densely connected convolutional networks, *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.2261-2269, 2017.
[12] F. Chollet, Xception: Deep learning with depthwise separable convolutions, *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.1800-1807, 2017.
[13] H. Müller, W. Müller, D. M. Squire, S. Marchand-Maillet and T. Pun, Performance evaluation in content-based image retrieval: Overview and proposals, *Pattern Recognit. Lett.*, vol.22, no.5, pp.593-601, 2001.
[14] E. O. Molina-Molina, S. Solorza-Calderón and J. Álvarez-Borrego, Classification of dermoscopy skin lesion color-images using fractal-deep learning features, *Appl. Sci.*, vol.10, no.17, 5954, 2020.
[15] M. A. Kassem, K. M. Hosny and M. M. Fouad, Skin lesions classification into eight classes for ISIC 2019 using deep convolutional neural network and transfer learning, *IEEE Access*, vol.8, pp.114822-114832, 2020.
[16] Z. Liu, R. Xiong and T. Jiang, CI-Net: Clinical-inspired network for automated skin lesion recognition, *IEEE Trans. Med. Imaging*, vol.42, no.3, pp.619-632, 2023.
[17] Cauvery, Siddalingaswamy, S. Pathan and N. D'souza, A multiclass skin lesion classification approach using transfer learning based convolutional neural network, *2021 7th International Conference on Bio Signals, Images, and Instrumentation (ICBSII)*, pp.1-6, 2021.