African Journal of
Management Engineering
and Technology

Afr. J. Manag. Engin. Technol., 2024, Vol. 2, N° 1, 11-23

Transfer Learning Approach Leveraging Efficient NetV2L to Enhance Skin Disease Prediction through Data Augmentation

Hamida S. [1, 2]*, El-Gannour O. [3, 4], Cherradi B. [3, 5]

[1] 2IACS Laboratory, ENSET of Mohammedia, Hassan II University of Casablanca, Mohammedia, Morocco.
[2] GENIUS Laboratory, SupMTI of Rabat Rabat, Morocco.
[3] EEIS Laboratory, ENSET of Mohammedia, Hassan II University of Casablanca Mohammedia, Morocco.
[4] LIASSE Laboratory, ENSA of Fez, Sidi Mohamed Ben Abdellah University, Fez, Morocco.
[5] CRMEF Casablanca-Settat, provincial section of El Jadida, 24000, El Jadida, Morocco.
**Corresponding author, Email address:* hamida@enset-media.ac.ma

Received *15 Jan 2024,*
Revised *21 Feb 2024,*
Accepted *25 Feb 2024*

Keywords:
- ✓ *Skin disease;*
- ✓ *Transfer learning;*
- ✓ *Augmentation;*
- ✓ *Data balancing;*
- ✓ *Performance*

Citation*: Hamida S., El-Gannour O., Cherradi B. (2024) Transfer Learning Approach Leveraging Efficient NetV2L to Enhance Skin Disease Prediction through Data Augmentation, Afr. J. Manag. Engg. Technol., 2(1), 11-23*

Abstract: Data imbalance is a common challenge in machine learning, particularly in medical image analysis tasks such as skin disease prediction. Melanoma, a type of skin cancer, is a prime example of a rare disease where the number of positive cases (minority class) is significantly smaller than the negative cases (majority class). This imbalance can lead to biased models that perform poorly in predicting rare diseases. Data augmentation techniques offer a solution by artificially increasing the number of minority class samples, thereby addressing the imbalance issue and improving the generalization of transfer learning models. In this paper, we focus on the application of data augmentation to balance the dataset for skin disease prediction. We specifically evaluate the performance of the Tuned EfficientNetV2L based classifier, a state-of-the-art model known for its efficiency and effectiveness in image classification tasks. Our experiments are conducted on a comprehensive collection of medical images and associated data of skin lesions sourced from various sources. These images represent diverse skin conditions, including both common and rare diseases, to ensure the robustness of our evaluation. To assess the performance of our approach, we employ various performance evaluation metrics such as accuracy, precision, and recall. These metrics provide insights into the classifier's ability to correctly classify skin lesions, especially rare diseases like melanoma.

1. Introduction

In medical diagnostics, as in numerous other domains, Machine Learning (ML) methods have brought about a paradigm shift. With the aim of facilitating early diagnosis and treatment of skin diseases, ML models analyze medical images in a critical manner. But a notable obstacle in the field of machine learning pertains to unbalanced datasets, which consist of significantly more instances of one class (e.g., healthy skin) than another (e.g., skin disease). An example of this is melanoma, a form of skin cancer, which illustrates how biased models may be when attempting to detect uncommon but critical cases (Khorshidi & Aickelin, 2021).

Balancing methods and data augmentation are implemented in order to tackle this concern. In order to alleviate the consequences of imbalance, balancing techniques are implemented in datasets. These techniques include over-sampling, under-sampling, and synthetic data generation (Krawczyk, 2016). The acquisition of knowledge from both majority and minority classes, thereby enhancing the overall performance of ML models, is contingent upon the application of these techniques (Dalianis, 2018).

In contrast, data augmentation is a process that transforms existing samples in order to increase the diversity of the dataset. This method facilitates the creation of a larger and more diverse training dataset (Luque *et al.,* 2019), which is especially advantageous when the available dataset is limited. Alternating the model's capability to generalize to unseen data can be achieved through the use of common data augmentation techniques, such as scaling, rotation, flipping, and noise addition to images (Shorten & Khoshgoftaar, 2019).

The limitations of data augmentation must, nevertheless, be taken into account. An overreliance on specific transformations may result in overfitting (Kim *et al.,* 2021), and the augmented data might not consistently reflect real-world fluctuations precisely. To ensure that the augmented dataset remains beneficial and representative for model training, it is therefore critical that data augmentation techniques be selected and implemented with extreme care (Maharana *et al.,* 2022).

We aim to enhance the performance of Transfer Learning (TL) models utilized for predicting skin diseases by examining the effects of data augmentation and balancing. TL is ideally suited for medical imaging tasks with limited data availability due to its utilization of pre-trained models expanded on large datasets and refined on smaller, domain-specific datasets (Daanouni *et al.,* 2020). Utilizing a dataset consisting of a variety of dermatoscopic images and demographic information, we aim to develop an effective TL model capable of accurately classifying patients with skin diseases. By expanding the dataset in both diversity and size, we intend to assess the efficacy of data augmentation in improving the performance of TL models (Mahjoubi *et al.,* 2023; Ouhmida *et al.,* 2021).

The paper is organized in the subsequent manner: A review of pertinent studies concerning the prediction of skin diseases and machine learning techniques is presented in Section II. The sources of medical images and demographic information that comprised the dataset utilized in this study are detailed in Section III. Methods utilized for data preprocessing, model training, and evaluation are also described. The performance metrics utilized to evaluate the efficacy of data augmentation and balancing are detailed in Section IV, which also contains the outcomes of our experiments. Concluding our study, Section V proposes avenues for further research in this field and discusses the implications of our findings.

2. Literature survey

A number of studies that have been conducted in the field of Data Balancing (DB) through the use of data augmentation have concentrated on improving the performance of Transfer Learning (TL) for the prediction of skin diseases. The use of data augmentation was suggested in the research paper (Mungloo-Dilmohamud *et al.,* 2022) as a means of achieving data balance and enhancing the performance of TL for the categorization of diabetic retinopathy. According to the findings of the study, methods such as rotation, flipping, and scaling have the potential to improve the generalizability of TL models. This is particularly advantageous for datasets that are either small or imbalanced.

The classification of skin lesions was the subject of yet another study (Shen *et al.,* 2022), which presented a novel data augmentation strategy that utilized deep learning. The performance of deep learning-based skin lesion categorization was improved by using this method, which incorporated affine transformation, color jittering, and histogram equalization. Due to the fact that it did not call for any new data collecting and was simple to put into action, the strategy could be considered cost-effective. The usefulness of this method could be investigated in subsequent study using a variety of medical imaging datasets and skin types.

Through the utilization of deep learning and the sparrow search algorithm, a novel method for the diagnosis of skin cancer was reported in a study (Balaha & Hassan, 2023). A high level of diagnostic accuracy and computing efficiency was reached through the use of this method. According to the findings, the sparrow search algorithm had the potential to optimize the parameters of the deep TL model, hence improving the level of performance and efficiency of the model in comparison to more conventional approaches.

For the purpose of categorizing skin lesions based on dermoscopic pictures, the authors of (Shetty *et al.*, 2022) suggested a method that combines Machine Learning (ML) with Convolutional Neural Network (CNN). A high level of classification accuracy was achieved by the utilization of CNN as a feature extractor and machine learning algorithms as classifiers, as evidenced by the study, which outperformed conventional approaches. CNN was utilized to successfully extract features from dermoscopic images, which resulted in an improvement in the performance of the model. It is possible that in the future, research may be conducted to evaluate its efficiency on various skin types and other medical imaging datasets.

Deep convolutional neural networks (CNNs) with TL models were used in another study (Ali *et al., 2021)* to propose an improved method for the categorization of skin cancer. The performance of the models was improved through the utilization of this strategy, which made use of the information acquired by pre-trained models and fine-tuned them for certain tasks.

The findings of these research collectively demonstrate that data augmentation approaches are beneficial in enhancing the performance of deep learning models for the categorization of skin lesions. Additionally, they demonstrate that the combination of TL and data augmentation is capable of effectively addressing the difficulty of dealing with tiny datasets through the process of medical picture analysis.

3. Methodology

3.1 Dataset

There is little doubt that the HAM10000 dataset is among the most complete collections of dermatoscopic images that have ever been gathered. It includes more than ten thousand pictures of skin lesions that have been tagged, and it was collected from more than seven thousand patients who remained anonymous. The ability to accurately identify and diagnose skin illnesses is made possible by this dataset, which presents a groundbreaking possibility (Tschandl *et al.*, 2018). In this part of the article, we will talk about the process that was used to acquire this dataset, as well as how it will be utilized in the field of medicine to offer support and insight for the diagnosis of skin illnesses.

For the purpose of skin cancer research, the Human Against Machine with 10000 training images dataset, often known as the HAM10000 dataset, has shown to be an instrumental and innovative resource. The authors of the study (Diwan *et al.*, 2023) carried out a comprehensive examination of the dataset in order to investigate the possibilities that it possesses. According to their findings, the HAM10000 dataset offers researchers an unparalleled quantity of data to work with, hence providing them with a comprehensive collection of photos that may be utilized for the purpose of skin cancer research. In addition, the authors discovered that the dataset possesses a high degree of accuracy and dependability. This is due to the fact that it includes photos derived from a variety of sources and is accompanied by the diagnosis that corresponds to each image. Researchers will find it much simpler to precisely diagnose and categorize incidences of skin cancer as a result of this. Some of the samples that were taken from the HAM10000 dataset are displayed in **Figure 1**.

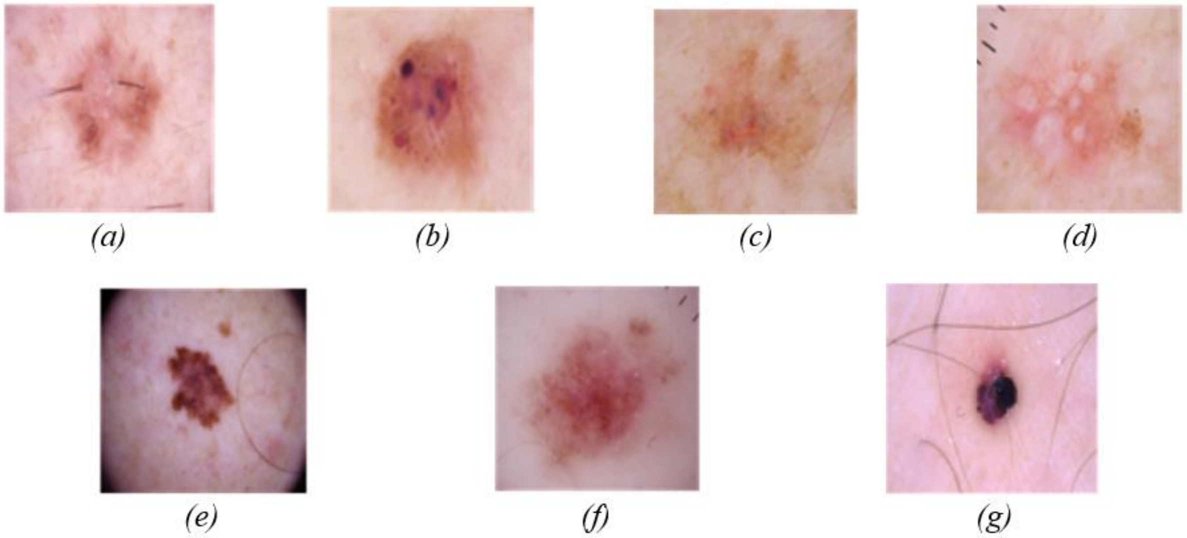


Figure 1. The HAM10000 dataset allowed for the extraction of seven different skin disease classifications as samples. (a) Actinic keratosis. (b) cancer of the basal cell. (c) Keratosis that is not harmful. (d) Dermatofibroma is the kind. (e) Melanoma. (f) Nevi that are melanocytic. (g) Lesions of the vasculature.

3.2 Data Pre-processing

The purpose of data pre-processing is to improve datasets by deleting features that are redundant and addressing missing values, while at the same time maintaining key features and evaluating the influence those features have on the results (Hamida *et al.*, 2023). In order to accomplish these goals, a number of pre-processing operations were carried out, such as feature encoding, the removal of redundant data, and the elimination of missing data (Nawi *et al.*, 2013). It is necessary for categorical variables to be encoded into numerical values in order for the majority of algorithms to be processed. The performance of the model is improved when categorical variables are encoded correctly since this enables the model to effectively understand and derive significant insights (Smolinska *et al.*, 2014).

3.2 Enhancement and Harmonization Methodology

An effective method for increasing the size of datasets is known as picture data augmentation (Shorten & Khoshgoftaar, 2019). This method involves the generation of modified copies of photographs. This method is especially useful when working with a limited amount of data because it can reduce the likelihood of overfitting and improve the capabilities of the model to adapt to new circumstances. Techniques like as flipping, rotating, shifting, shearing, zooming, cropping, adjusting colors, and modifying brightness are examples of common augmentation techniques. It is possible for data augmentation to either supplement or replace strategies for balancing (Abdi & Hashemi, 2016).

Altering datasets in a basic manner can be accomplished by flipping photos either horizontally or vertically. It is possible to rotate images around an axis by any degree between 1 and 359, around the center, or around another point. This is made possible using rotation. However, excessive rotation may lead to labels that are not consistent with one another. Shearing is a technique that moves a portion of a picture in the direction of a parallelogram, so changing its orientation (Al-masni *et al.*, 2020). Shifting is a technique that includes moving all of the pixels of an image to a different place. Through the process of zooming, images are produced with varying degrees of zoom by randomly adding additional pixels. Cropping is the process of deleting a piece of an image, which can be done in a random manner or by center-cropping, which is done when the center of the image has more information than the corners. Color adjustment is used to change the values of individual pixels, whereas brightness modification is used to change the brightness of the image as a whole. In order to equalize the categories, Data Balancing (DB) is implemented prior to the training phase. This is done since the number of images that are assigned to each category is inconsistent. In **Table 1**, the parameters that were utilized for the data augmentation technique are presented.

Table 1. Data augmentation parameters that are applied in the process of carrying out the procedure.

Parameter	Value
Rotation	20°
Shifting the width and height	20%
Shearing	25%
Flipping	horizontal
Brightness range	[0.5 : 1.5]

As a result of doing an analysis of the characteristics of the dataset, it became apparent that the distribution of skin disorders in humans displays very different patterns. Melanocytic nevi were found to be the condition that was most widespread among individuals, while Dermatofibroma was found to be the disease that manifested itself the least frequently. The significance of these findings lies in the fact that they point to the existence of an imbalanced dataset, in which the target variable has a different number of observations for each class. Two different approaches, namely class weights and data augmentation, were utilized in this research project in order to address this problem.

One of the strategies involves the utilization of class weights, which are introduced into the model during the training process in order to improve the model's sensitivity to classes that are less abundant in the dataset. There is an inverse relationship between the number of samples that belong to a particular class and the weighting value that is allocated to that class. As a result, classes that have a smaller number of samples are given greater weights in order to guarantee that they receive more attention during the training of the model. The suggested model is constructed using class weights, which are outlined in **Table 2**, which provides an overview of these weights.

Table 2. Class Coefficients Were Employed in Order to Complement the HAM10000 Dataset Classes.

Reference	Classes	Coefficients
0	Actinic keratoses	**4.0**
1	Basal cell carcinoma	**4.0**
2	Benign keratosis-like lesions	**2.0**
3	Dermatofibroma	**7.0**
4	Melanoma	**4.0**
5	Melanocytic nevi	**2.0**
6	Vascular lesions	**8.0**

3.3. *Transfer Learning Model*

Transfer learning, often known as TL, is a method of machine learning in which a model that has been trained on one task is applied to another task that is not connected to the first job. This strategy makes use of features that have been learnt in the past, hence minimizing the amount of data and computational resources that are required to train a new model (Gannour et al., 2022). Through its application in a variety of fields, including computer vision and natural language processing, TL has demonstrated its effectiveness. The use of a pre-trained model as a feature extractor is a frequent approach. In this method, the lower layers extract features from new data, and then a new classifier is trained on these features. The performance of this strategy has been shown to improve across a wide variety of tasks, particularly in situations where there is a limited amount of labeled data available for a new task (Srinivasu et al., 2021).

EfficientNetV2 models are a collection of convolutional neural networks (CNNs) that were developed with the intention of improving the velocity of training and the efficiency of the parameters. The optimization of the network's architecture and hyperparameters is accomplished by the utilization of a scaling method and a neural architecture search algorithm both simultaneously. EfficientNetV2 models are designed specifically for image classification tasks and are offered in a variety of scales, each of which corresponds to a distinct model architecture and number of parameters. This results in a greater number of parameters and a higher level of accuracy. The architecture search space has been broadened to incorporate additional operations such as Fused-MBConv, in addition to the conventional CNN techniques that were previously available. An illustration of the structure of MBConv and Fused-MBConv is found in **Figure 2**.

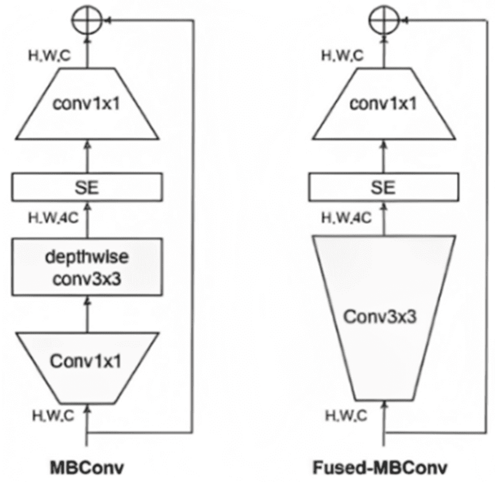


Figure 2. MBConv and Fused-MBConv Structure.

The EfficientNetV2 models are an extension of the EfficientNetV1 models, and they have showed performances that are state-of-the-art across a variety of vision tasks. These tasks include picture classification, object identification, and semantic segmentation. The efficiency with which these models do computations is well-known knowledge.

3.4. Performance Metrics

When determining how accurate a model's predictions are, performance metrics are an extremely important factor to consider (Lamalem, Hamida, Housni, et al., 2022; Lamalem, Hamida, Tazouti, et al., 2022). The effectiveness of a model's predictions can be evaluated with their help, as can the comparison of other models, the identification of areas that could be improved, and the direction of the development of new algorithms (Hamida et al., 2019). The concepts of accuracy, precision, recall, F1 score, and the ROC curve are among the most frequently used performance metrics in the field of machine learning. The percentage of accurate predictions that the model generates is what is meant by the term "accuracy." In order to determine whether or not a model is able to avoid producing false positives, precision is calculated by determining the percentage of true positives among all positive predictions. The ability of the model to capture all positive cases is indicated by the recall, which is a measurement that determines the percentage of genuine positives found among all actual positives. Precision and recall are also components of the F1 score, which represents a well-rounded evaluation of the entire performance of the model. The Receiver Operating Characteristic (ROC) curve is a graphical depiction of the performance of the model, which is helpful for comparing various models.

Besides these measurements, there are more metrics that can be utilized to evaluate the success of a machine learning model. These metrics include log loss, AUC (Area Under the Curve), and Matthews correlation coefficient. The difference between the projected likelihood of an event and the actual probability of doing so is what is meant to be measured by log loss. Area under the ROC curve (AUC) is a metric that is utilized for the purpose of comparing different models. When comparing projected values to actual values, the Matthews correlation coefficient is used to evaluate the degree of correlation (Fürnkranz et al., 2011). When conducting an evaluation of the performance of a machine learning model, it is imperative to take into consideration all of these measures. Every statistic provides a different point of view on the performance of the model and has the potential to indicate areas that could use some improvement. Another thing that needs to be taken into consideration is the context, which includes things like the particular application of the model (Biggerstaff, 2008).

4. Experimental results

The Cross-validation approach is utilized in this part for the purpose of assessing the performance of both the baseline and tuned TL models. In two separate tests, we evaluate the quality of the model by employing a variety of metrics: The first experiment consisted of maintaining the baseline model setting and making use of the HAM10000 dataset with preprocessing. During the second experiment, the EfficientNetV2 pre-trained model was fine-tuned using the enhanced dataset that was derived from the HAM10000 dataset. This was done with preprocessing.

4.1. Establishing a baseline model with the default dataset

In terms of performance, the results of the baseline pre-trained model on the HAM10000 dataset are extremely remarkable. To guarantee that the findings would be comparable, the model was trained for a total of 25 epochs using the same configuration. We have displayed the learning and validation curves for the baseline pre-trained model in order to guarantee that we will not experience any problems related to overfitting or underfitting. The training and validation curves are plotted in **Figure 3**, which depicts the accuracy and loss graphs of the model that has been pre-trained.

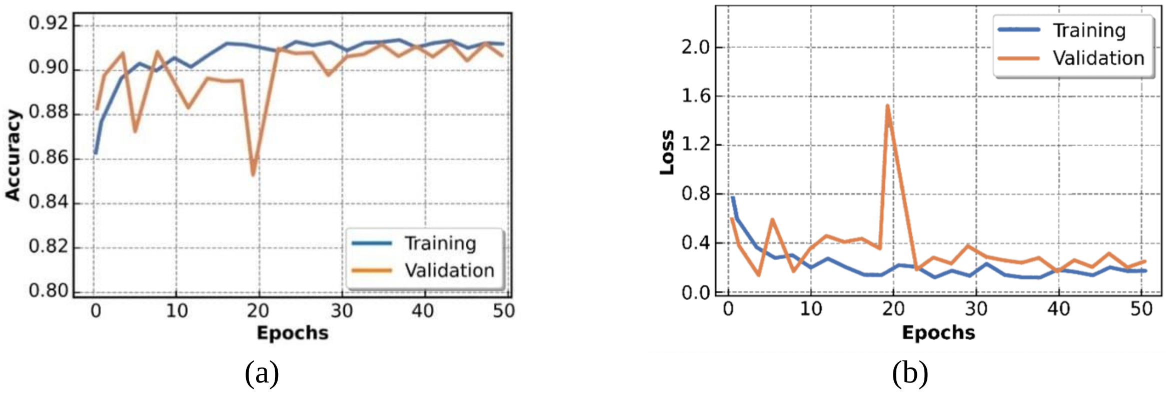
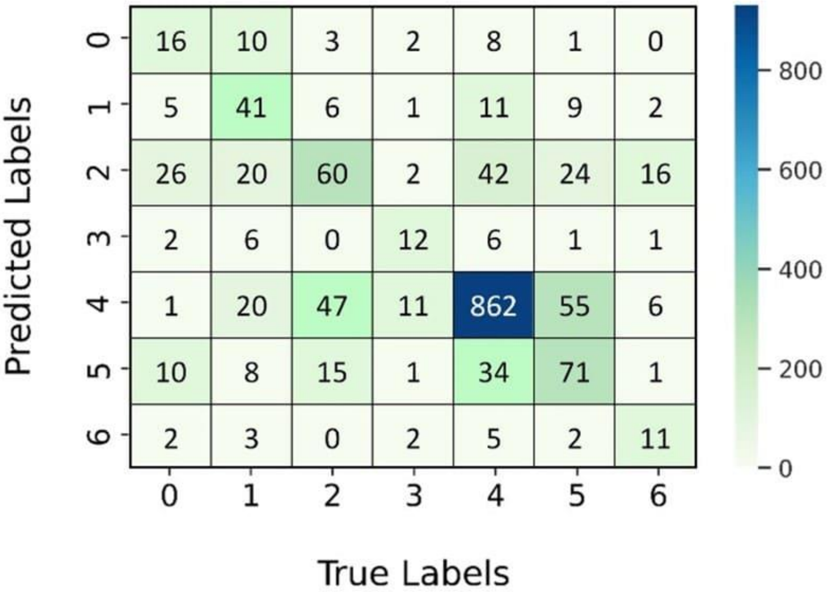


Figure 3. Accuracy and Loss curves obtained by training and validating the baseline model. (a) Accuracy training curve. (b) Loss training curve.

Based on the findings, it appears that the model is able to correctly categorize the photographs contained in the dataset. This is demonstrated by the accuracy and loss graphs, which show that the pre-trained models successfully categorize the images contained in the dataset, obtaining a high level of accuracy of approximately 91%. It is also clear from the loss graphs that the models perform admirably when it comes to the classification of the photos, with the lowest loss being at around 0.3.

Additional evidence that the models are capable of successfully classifying the photos contained in the dataset is provided by the learning and validation curves. The validation curves demonstrate that the models are able to generalize quite well, while the learning curves demonstrate that the models are able to learn from the data in an efficient manner.

In point of fact, the training outcomes of the baseline model using the default dataset are rather remarkable. Eighty percent of the HAM10000 dataset was used for training the model, and twenty-five percent was used for testing. Calculations were made to determine the outcomes of the multi-class categorization, and the findings are shown in **Figure 4**.



	0	1	2	3	4	5	6
0	16	10	3	2	8	1	0
1	5	41	6	1	11	9	2
2	26	20	60	2	42	24	16
3	2	6	0	12	6	1	1
4	1	20	47	11	862	55	6
5	10	8	15	1	34	71	1
6	2	3	0	2	5	2	11

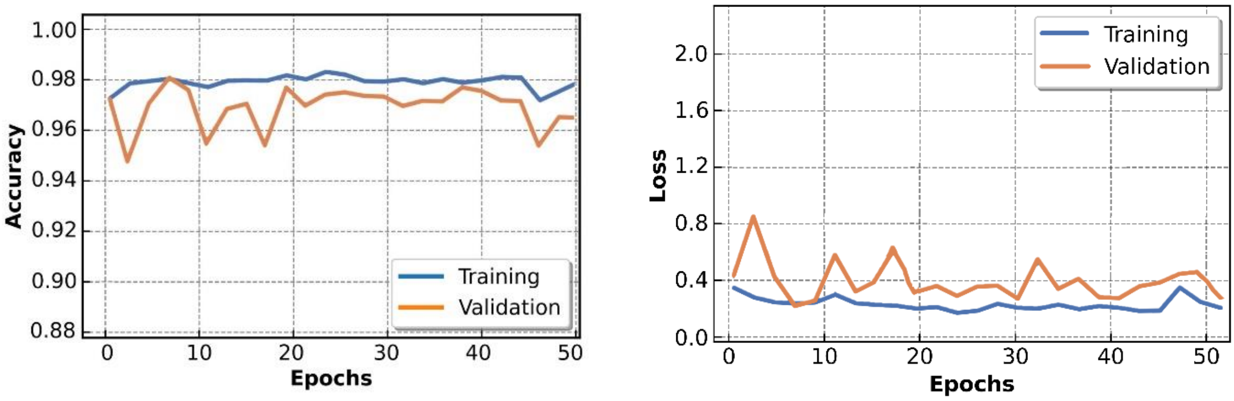
Figure 4. Confusion matrix representing the performance of the baseline classifier.

The examples that are considered to be True Positives (TP) are those in which the prediction is accurately predicted and the actual value is also accurate. It may be deduced from this that the model was successful in making correct predictions for the positive cases. Those instances that are referred to as False Positives (FP) are those in which the forecast is positive but the actual result is negative. The fact that the model was able to make inaccurate predictions for some circumstances is demonstrated by this. The cases that are considered to be True Negative (TN) are those in which the forecast predicts a negative value and the actual value is also negative. It may be deduced from this that the model was successful in precisely predicting the cases that were negative. The examples that are considered to be False Negatives (FN) are those in which the prediction is negative but the actual result obtained is positive. It may be deduced from this that the model was not able to accurately forecast a few of the positive cases encountered.

4.2. TL model with expanded dataset

An accurate classification of the images contained in the expanded HAM10000 dataset may be achieved by the pre-trained models, as demonstrated by the subsequent findings. When it comes to understanding the performance of a machine learning model, the training and validation curves are, without a doubt, an essential component. The curves can provide insight into the effective training of the model, probable situations of under-fitting or over-fitting, and the number of epochs that are required to obtain convergence. This is accomplished by displaying the epochs on the x-axis and improvement on the y-axis. For the purpose of gaining a deeper understanding of this idea, the purpose of this essay is to analyze the ways in which the correlations that exist between the training curve and the validation curve can be utilized to evaluate ideal performance.

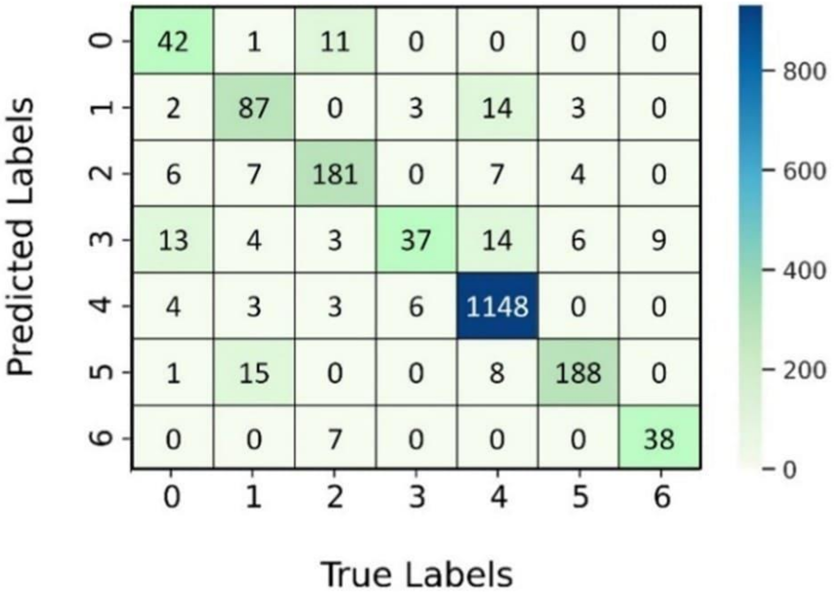
The accuracy of the model is more than that of the baseline model, and the loss is less than the baseline model. The fact that the pre-trained models are able to learn the characteristics of the dataset more effectively than the baseline model is demonstrated by this example. As an additional point of interest, the learning and validation curves demonstrate that the model does not exhibit either overfitting or underfitting, which is a significant factor in ensuring that the model is accurate. by drawing the training and validation curves, **Figure 5** displays the accuracy and loss graphs of the pre-trained model based on the supplemented data. These graphs were created by using the augmented data.



(a) (b)

Figure 5. Accuracy and Loss curves obtained by training and validating the pre-trained model based on the augmented data. (a) Accuracy training curve. (b) Loss training curve.

In light of the matrix that is displayed in **Figure 6**, it is evident that the classification of the pre-trained model that is based on balanced data is improved when compared with the outcomes of an earlier experiment.



	0	1	2	3	4	5	6
0	42	1	11	0	0	0	0
1	2	87	0	3	14	3	0
2	6	7	181	0	7	4	0
3	13	4	3	37	14	6	9
4	4	3	3	6	1148	0	0
5	1	15	0	0	8	188	0
6	0	0	7	0	0	0	38

Figure 6. Confusion matrix representing the performance of the pre-trained classifier based on balanced dataset.

4.3. Discussion

When it comes to improving the performance of TL models for skin disease prediction, data balance through data augmentation is an extremely important factor by which to consider. The capacity of models to exploit knowledge obtained from one task to boost performance on a related task is made possible by TL. This is especially useful in the field of medical image classification, where the supply of data may be restricted. When it comes to the prediction of skin diseases using TL, one of the most significant challenges is class imbalance. This occurs when the quantity of samples in each class varies, which results in a bias towards the class that is in the majority. It is possible that this bias will have a negative impact on the performance of the model when applied to minority classes, which may not have sufficient training instances present. Class imbalance can be addressed through the use of data augmentation, which involves boosting the number of samples that belong to minority classes by utilizing transformations such as rotation, scaling, flipping, and adding noise to pictures that already contain minority class members. By increasing the number of samples from minority classes, data augmentation helps to achieve a more equitable distribution of classes and enhances the model's capacity to generalize to new examples.

For the purpose of this investigation, we conducted two experiments with a dataset consisting of skin lesion photos in order to assess the efficacy of data augmentation for classification models. Following the division of the dataset into a training set and a test set, the training set was enhanced by the use of a variety of strategies. After the pre-trained model of EfficientNetV2 was applied to the expanded dataset, it was then fine-tuned. The results of the second experiment demonstrated that the performance of the model on new cases was greatly improved following the implementation of data augmentation. When compared to the model that was trained without the use of data augmentation, the algorithm that was trained with data augmentation achieved a better level of accuracy on the test set. When it comes to increasing the efficacy of TL for skin disease prediction. Additionally included in **Table 3** are the measures that pertain to the evaluation of performance models.

Table 3. Performance Evaluation Metric of The Baseline and Fine-Tuned Model.

Models	**Accuracy**	**Precision**	**Sensitivity**	**Specificity**	**Loss**
Baseline model (Unbalanced data)	91,87%	49,98%	93,66%	44,37%	9.03%
Fine-tuned model (Balanced data)	97,65%	80,09%	98,23%	82,23%	4.32%

The utilization of TL models for the detection and classification of skin conditions does, however, have some limitations. Because of the slow speed of the boosting method and the huge number of features that are used, the process of training the classifier takes a significant amount of time.

Conclusion

In conclusion, we have proposed a way to produce synthetic data by utilizing picture augmentation in order to achieve a more balanced dataset and improve the performance of a TL model for skin lesion classification. The accuracy of our strategy was found to be 6% higher than that of the baseline model when it was tested on the HAM10000 dataset to see how well it performed.

In the work that we intend to do in the future, we intend to integrate data augmentation with other methods, such as feature selection or fine-tuning pre-trained models, in order to further increase performance. In addition, we intend to apply this methodology to additional medical imaging datasets, such as X-ray or MRI, in order to evaluate how effective, it is in enhancing performance in a variety of medical specialties.

Disclosure statement: *Conflict of Interest:* The authors declare that there are no conflicts of interest.
Compliance with Ethical Standards: This article does not contain any studies involving human or animal subjects.

References

Abdi, L., & Hashemi, S. (2016). To Combat Multi-Class Imbalanced Problems by Means of Over-Sampling Techniques. IEEE Transactions on Knowledge and Data Engineering, 28(1), 238–251. https://doi.org/10.1109/TKDE.2015.2458858

Ali, M. S., Miah, M. S., Haque, J., Rahman, M. M., & Islam, M. K. (2021). An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models. Machine Learning with Applications, 5, 100036. https://doi.org/10.1016/j.mlwa.2021.100036

Al-masni, M. A., Kim, D.-H., & Kim, T.-S. (2020). Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification. Computer Methods and Programs in Biomedicine, 190, 105351. https://doi.org/10.1016/j.cmpb.2020.105351

Balaha, H. M., & Hassan, A. E.-S. (2023). Skin cancer diagnosis based on deep transfer learning and sparrow search algorithm. Neural Computing and Applications, 35(1), 815–853. https://doi.org/10.1007/s00521-022-07762-9

Biggerstaff, B. J. (2008). Confidence intervals for the difference of two proportions estimated from pooled samples. Journal of Agricultural, Biological, and Environmental Statistics, 13(4), 478–496. https://doi.org/10.1198/108571108X379055

Daanouni, O., Cherradi, B., & Tmiri, A. (2020). Diabetes Diseases Prediction Using Supervised Machine Learning and Neighbourhood Components Analysis. Proceedings of the 3rd International Conference on Networking, Information Systems & Security, 1–5. https://doi.org/10.1145/3386723.3387887

Dalianis, H. (2018). Evaluation Metrics and Evaluation. In H. Dalianis, Clinical Text Mining (pp. 45–53). Springer International Publishing. https://doi.org/10.1007/978-3-319-78503-5_6

Diwan, T., Shukla, R., Ghuse, E., & Tembhurne, J. V. (2023). Model hybridization & learning rate annealing for skin cancer detection. Multimedia Tools and Applications, 82(2), 2369–2392. https://doi.org/10.1007/s11042-022-12633-5

Fürnkranz, J., Chan, P. K., Craw, S., Sammut, C., Uther, W., Ratnaparkhi, A., Jin, X., Han, J., Yang, Y., Morik, K., Dorigo, M., Birattari, M., Stützle, T., Brazdil, P., Vilalta, R., Giraud-Carrier, C., Soares, C., Rissanen, J., Baxter, R. A., … De Raedt, L. (2011). Model Evaluation. In C. Sammut & G. I. Webb (Eds.), Encyclopedia of Machine Learning (pp. 683–683). Springer US. https://doi.org/10.1007/978-0-387-30164-8_550

Gannour, O. E., Hamida, S., Saleh, S., Lamalem, Y., Cherradi, B., & Raihani, A. (2022). COVID-19 Detection on X-Ray Images using a Combining Mechanism of Pre-trained CNNs. International Journal of Advanced Computer Science and Applications, 13(6). https://doi.org/10.14569/IJACSA.2022.0130668

Hamida, S., Cherradi, B., El Gannour, O., Raihani, A., & Ouajji, H. (2023). Cursive Arabic handwritten word recognition system using majority voting and k-NN for feature descriptor selection. Multimedia Tools and Applications. https://doi.org/10.1007/s11042-023-15167-6

Hamida, S., Cherradi, B., Raihani, A., & Ouajji, H. (2019). Performance Evaluation of Machine Learning Algorithms in Handwritten Digits Recognition. 2019 1st International Conference on Smart Systems and Data Science (ICSSD), 1–6. https://doi.org/10.1109/ICSSD47982.2019.9003052

Khorshidi, H. A., & Aickelin, U. (2021). Constructing classifiers for imbalanced data using diversity optimisation. Information Sciences, 565, 1–16. https://doi.org/10.1016/j.ins.2021.02.069

Kim, Y., Kim, Y., Yang, C., Park, K., Gu, G. X., & Ryu, S. (2021). Deep learning framework for material design space exploration using active transfer learning and data augmentation. Npj Computational

Materials, 7(1), 140. https://doi.org/10.1038/s41524-021-00609-2

Krawczyk, B. (2016). Learning from imbalanced data: Open challenges and future directions. Progress in Artificial Intelligence, 5(4), 221–232. https://doi.org/10.1007/s13748-016-0094-0

Lamalem, Y., Hamida, S., Housni, K., Ouhmida, A., & Cherradi, B. (2022). Evaluating Systems Reliability With A New Method Based on Node Cutset. 2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), 1–4. https://doi.org/10.1109/IRASET52964.2022.9737956

Lamalem, Y., Hamida, S., Tazouti, Y., Gannour, O. E., Housni, K., & Cherradi, B. (2022). Evaluating multi-state systems reliability with a new improved method. Bulletin of Electrical Engineering and Informatics, 11(3), 1568–1576. https://doi.org/10.11591/eei.v11i3.3509

Luque, A., Carrasco, A., Martín, A., & de las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. Pattern Recognition, 91, 216–231. https://doi.org/10.1016/j.patcog.2019.02.023

Maharana, K., Mondal, S., & Nemade, B. (2022). A review: Data pre-processing and data augmentation techniques. Global Transitions Proceedings, 3(1), 91–99. https://doi.org/10.1016/j.gltp.2022.04.020

Mahjoubi, M. A., Hamida, S., Gannour, O. E., Cherradi, B., Abbassi, A. E., & Raihani, A. (2023). Improved Multiclass Brain Tumor Detection using Convolutional Neural Networks and Magnetic Resonance Imaging. International Journal of Advanced Computer Science and Applications, 14(3). https://doi.org/10.14569/IJACSA.2023.0140346

Mungloo-Dilmohamud, Z., Heenaye-Mamode Khan, M., Jhumka, K., Beedassy, B. N., Mungloo, N. Z., & Peña-Reyes, C. (2022). Balancing Data through Data Augmentation Improves the Generality of Transfer Learning for Diabetic Retinopathy Classification. Applied Sciences, 12(11), 5363. https://doi.org/10.3390/app12115363

Nawi, N. M., Atomi, W. H., & Rehman, M. Z. (2013). The Effect of Data Pre-processing on Optimized Training of Artificial Neural Networks. Procedia Technology, 11, 32–39. https://doi.org/10.1016/j.protcy.2013.12.159

Ouhmida, A., Raihani, A., Cherradi, B., & Terrada, O. (2021). A Novel Approach for Parkinson's Disease Detection Based on Voice Classification and Features Selection Techniques. International Journal of Online and Biomedical Engineering (iJOE), 17(10), 111. https://doi.org/10.3991/ijoe.v17i10.24499

Shen, S., Xu, M., Zhang, F., Shao, P., Liu, H., Xu, L., Zhang, C., Liu, P., Zhang, Z., Yao, P., & Xu, R. X. (2022). A Low-Cost High-Performance Data Augmentation for Deep Learning-Based Skin Lesion Classification. BME Frontiers, 2022, 2022/9765307. https://doi.org/10.34133/2022/9765307

Shetty, B., Fernandes, R., Rodrigues, A. P., Chengoden, R., Bhattacharya, S., & Lakshmanna, K. (2022). Skin lesion classification of dermoscopic images using machine learning and convolutional neural network. Scientific Reports, 12(1), 18134. https://doi.org/10.1038/s41598-022-22644-9

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. Journal of Big Data, 6(1), 60. https://doi.org/10.1186/s40537-019-0197-0

Smolinska, A., Hauschild, A.-C., Fijten, R. R. R., Dallinga, J. W., Baumbach, J., & van Schooten, F. J. (2014). Current breathomics—A review on data pre-processing techniques and machine learning in metabolomics breath analysis. Journal of Breath Research, 8(2), 027105. https://doi.org/10.1088/1752-7155/8/2/027105

Srinivasu, P. N., SivaSai, J. G., Ijaz, M. F., Bhoi, A. K., Kim, W., & Kang, J. J. (2021). Classification of Skin Disease Using Deep Learning Neural Networks with MobileNet V2 and LSTM. Sensors, 21(8), 2852. https://doi.org/10.3390/s21082852

Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific Data, 5(1), 180161. https://doi.org/10.1038/sdata.2018.161

(2024) ; https://revues.imist.ma/index.php/ajmet/index