Improving skin diseases prediction through data balancing via classes weighting and transfer learning

Oussama El Gannour[1], Soufiane Hamida[1,2], Yasser Lamalem[3], Mohamed Amine Mahjoubi[1], Bouchaib Cherradi[1,4], Abdelhadi Raihani[1]
[1]EEIS Laboratory, ENSET of Mohammedia, Hassan II University of Casablanca, Mohammedia, Morocco
[2]GENIUS Laboratory, SupMTI of Rabat, Rabat, Morocco
[3]Computer Research Laboratory (L@RI), Faculty of Science, University of Ibn Tofail, Kenitra, Morocco
[4]STIE Team, CRMEF Casablanca-Settat, Provincial Section of El Jadida, El Jadida, Morocco

Article Info

Article history:

Received Feb 12, 2023
Revised Jul 13, 2023
Accepted Sep 11, 2023

Keywords:

Classes weighting approach
Data balancing
EfficientNetV2L model
Medical diagnosis
Skin diseases
Transfer learning

ABSTRACT

Skin disease prediction using artificial intelligence has shown great potential in improving early diagnosis and treatment outcomes. However, the presence of class imbalance within skin disease datasets poses a significant challenge for accurate prediction, particularly for rare diseases. This study proposes a novel approach to address class imbalance through data balancing using classes weighting, coupled with transfer learning techniques, to enhance the performance of skin disease prediction models. Two experiments were conducted using a tuned EfficientNetV2L based classifier. In the first experiment, a default dataset structure was utilized for training and testing. The second experiment involved employing classes weighting approach to balance the dataset. The effectiveness of the proposed approach is evaluated using the ISIC 2018 dataset, which comprises a diverse collection of skin lesion images. By assigning appropriate weights to different classes based on their prevalence, the proposed method aims to balance the representation of rare disease classes. To evaluate the performance of the proposed methodology, several performance evaluation metrics, including accuracy, precision, and recall, were employed. These findings revealed that the balanced dataset achieved enhanced generalization, mitigating the biases associated with class imbalance. As a result, the efficacy of artificial intelligence models is enhanced.

Corresponding Author:

Oussama El Gannour
EEIS Laboratory, ENSET of Mohammedia, Hassan II University of Casablanca
Mohammedia, Morocco
Email: oussama.elgannour@gmail.com

1. INTRODUCTION

Skin diseases pose significant challenges in the field of medical diagnostics and treatment. With advancements in artificial intelligence (AI), there is a growing interest in leveraging AI models for accurate and efficient skin disease prediction [1], [2]. However, building effective AI models for skin disease prediction encounters a common obstacle: class imbalance within the datasets [3]. Class imbalance occurs when certain skin disease categories have a significantly higher representation than others [4]. This disparity can lead to biased models that perform poorly on underrepresented classes, limiting their effectiveness in clinical practice. To address this issue, data balancing techniques play a crucial role in improving the performance and generalization of AI models [5], [6]. Data balancing aims to equalize the representation of different classes within the dataset used for training [7], [8]. Various methods, such as oversampling, undersampling, and synthetic data generation, have been employed to alleviate class imbalance [9]. However, in the context of skin disease prediction [10], there is a need for effective data balancing techniques that specifically account for class imbalance and improve transfer learning performance [11]. In this article, we focus on the utilization of class weighting as a data balancing technique to enhance transfer learning performance for skin disease prediction [11]. Class weighting assigns higher weights to instances of underrepresented classes during model training, allowing the model to give more attention to these classes and mitigate the bias towards the majority class [12]. By considering the impact of class imbalance on model training, we aim to improve the accuracy and robustness of AI models for skin disease prediction, ultimately advancing the field of dermatology [13], [14].

The primary objective of this study is to investigate the efficacy of class weighting as a data balancing technique in the context of transfer learning for skin disease prediction. We aim to address the limitations of existing approaches by providing a comprehensive evaluation of the performance and generalization capabilities of AI models when class weighting is applied [15]. By conducting rigorous experiments and assessments, we seek to gain insights into the potential benefits and practical implications of class weighting for skin disease prediction. To evaluate the performance of the proposed approach, we utilize a diverse and comprehensive dataset consisting of medical images and associated data of skin lesions sourced from various clinical sources. The dataset encompasses a wide range of skin diseases, including both common and rare conditions. This diversity ensures that the evaluation captures the challenges associated with class imbalance, especially in the context of rare and underrepresented skin disease categories.

In our methodology, we employ transfer learning, a widely adopted approach in the field of AI, which leverages pre-trained models on large-scale datasets and adapts them for the specific task of skin disease prediction [16]. Transfer learning allows the model to benefit from the knowledge and representations learned from a different but related task, enabling efficient training and improved performance [17], [18]. However, the effectiveness of transfer learning can be hindered by class imbalance within the dataset, necessitating the application of data balancing techniques such as class weighting. To assess the effect of class weighting, we compare the performance of AI models trained with and without class weighting on various skin disease prediction tasks. We employ widely accepted evaluation metrics such as accuracy, precision, recall, and F1-score to assess the performance of the models [19]. Additionally, we analyze the effects of class weighting on the model's ability to correctly identify and classify different skin disease categories, paying particular attention to the improvement in predicting rare and underrepresented diseases.

This paper is organized as follows: section 2 provides a detailed explanation of the proposed methodology for balancing the dataset, including the utilization of classes weighting. Section 3 elucidates the structure of the dataset and the methods employed in this study. Section 4 showcases the experimental results and presents the performance evaluation of the proposed approach. Finally, section 5 concludes the paper by summarizing the findings and discussing potential avenues for future research.

2. PROPOSED DATASET BALANCING MECHANISM

The accurate prediction of skin diseases using AI models is hindered by the challenge of class imbalance within the datasets. In this section, we present our proposed methodology for balancing the dataset, specifically through the utilization of classes weighting. This technique aims to address the disparity in the representation of different skin disease categories, ultimately improving the transfer learning performance for skin disease prediction.

2.1. Classes weighting approach

Skin disease datasets commonly exhibit class imbalance, where certain disease categories are overrepresented while others are underrepresented. This imbalance can lead to biased model training, as the models tend to favor the majority class, resulting in poor prediction performance for the minority classes. To mitigate this issue, data balancing techniques are essential to create a more equitable distribution of samples across different classes [20].

In our proposed methodology, we employ classes weighting as a data balancing technique. Classes weighting assigns higher weights to instances of the underrepresented classes during the model training phase. By doing so, the model pays more attention to the minority classes, thereby reducing the bias towards the majority class and improving the overall performance and generalization capabilities of the AI model [21].

2.2. Implementation of classes weighting

To implement classes weighting, we adjust the loss function during the training process. The loss function is modified to give more importance to misclassifications in the minority classes. By assigning higher loss weights to these misclassifications, the model is encouraged to focus on correctly predicting the underrepresented classes, effectively reducing the negative impact of class imbalance on the learning process.

2.3. Determining class weights

The determination of class weights is a crucial aspect of the proposed methodology. The weights are typically calculated based on the inverse class frequencies and a proportional mechanism that ensures fair representation of all classes during training. This allows the model to learn from both the majority and minority classes, leading to improved performance on the underrepresented classes. In our study, we addressed the issue of class imbalance in the skin disease dataset by implementing class weighting using mathematical functions. We conducted an analysis of the dataset and identified the majority and minority classes. To calculate the class weights, we employed two mathematical equations: the inverse class frequency and proportional class weighting. For the inverse class frequency weight calculation, we divided the total number of instances by the number of instances in each class. The formula for calculating the inverse class frequency weight for i^{th} *class* is as (1):

$$FW_i = \frac{\sum_{k=0}^{n} I}{\sum I_i} \tag{1}$$

Where:
FW_i is the inverse class frequency weight for i^{th} *class.*
I is the total number of instances.
I_i is the total instances in i^{th} *class*.

Additionally, we utilized proportional class weighting to assign weights to each class proportionally to the number of classes in the datasets. The formula for calculating the proportional class weight for i^{th} *class* is given by:

$$W_i = log_{10}(FW_i) \times N \tag{2}$$

Where:
W_i is the weight for i^{th} *class*.
N is the number of classes in the dataset.

By assigning these calculated weights to the instances in the dataset, we ensured that instances from the minority class received higher weights. During the model training phase, we incorporated the assigned class weights into the loss function, enabling the model to prioritize the underrepresented classes. Table 1 presents the class weights employed in the construction of the planned model. In the next section, we delve into the structure of the dataset and the methods employed in our study. By combining the proposed classes weighting methodology with robust dataset structures and effective training techniques, we aim to enhance the transfer learning performance for skin disease prediction, ultimately advancing diagnostic accuracy, and patient care in the field of dermatology.

Table 1. Class weights utilized for model construction in skin disease prediction

Ref. number	Class	Weight
0	Actinic keratoses	9.0
1	Basal cell carcinoma	9.0
2	Benign keratosis-like lesions	7.0
3	Dermatofibroma	14.0
4	Melanoma	7.0
5	Melanocytic nevi	1.0
6	Vascular lesions	14.0

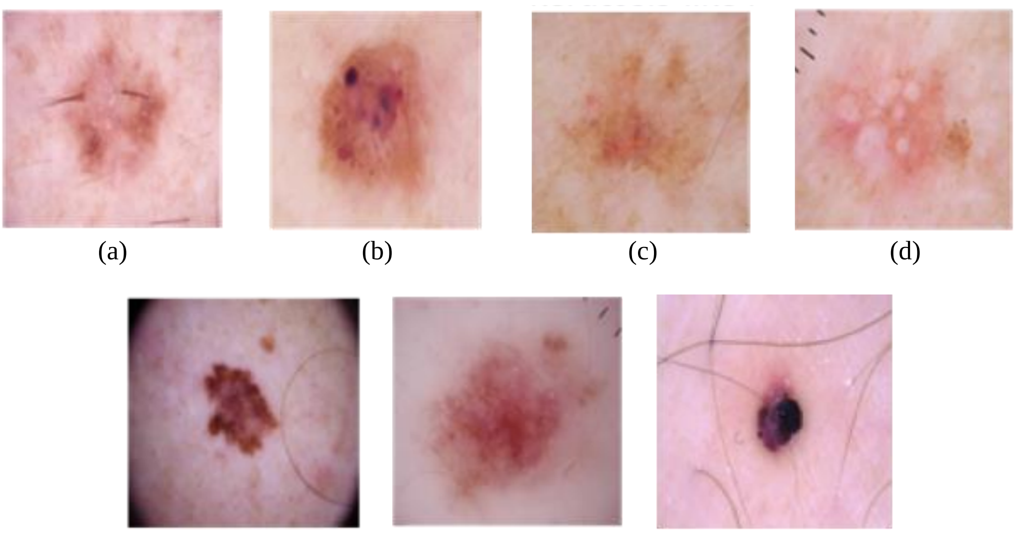
3. DATASET STRUCTURE AND METHOD

In this section, we provide an overview of the dataset structure and the methods employed in our study, with a particular focus on the utilization of classes weighting for improving transfer learning performance in skin disease prediction.

3.1. ISIC 2018 dataset: composition and characteristics

We begin by describing the dataset used in our research, which comprises a comprehensive collection of medical images and associated data of skin lesions sourced from various reliable sources. The dataset encompasses diverse skin disease categories, including common and rare conditions, providing a representative sample of the challenges faced in real-world skin disease prediction scenarios.

The ISIC 2018 dataset is a widely recognized benchmark dataset for skin disease analysis and consists of a diverse collection of dermoscopic images [22]. These images were acquired from various clinical sources and encompass a broad range of skin lesions, including malignant melanoma, basal cell carcinoma, benign nevi, and seborrheic keratosis. The dataset comprises a total of 10000 number of images, each labeled with a corresponding skin disease category. The images are captured using high-resolution dermoscopes, enabling the visualization of intricate features and patterns on the skin's surface. The dataset provides an invaluable resource for training and evaluating AI models in the field of skin disease prediction. In Figure 1, we present a selection of sample images extracted from the ISIC 2018 dataset. These images serve as visual representations of the diverse skin lesions included in the dataset. Each image showcases distinct visual characteristics and patterns associated with different skin disease categories, providing valuable insights into the complexity and variability of skin diseases captured within the dataset. The samples displayed in Figure 1 exhibit a range of skin conditions, including actinic keratosis in Figure 1(a), basal cell carcinoma in Figure 1(b), benign keratosis in Figure 1(c), dermatofibroma in Figure 1(d), melanoma in Figure 1(e), melanocytic nevi in Figure 1(f), and vascular lesions in Figure 1(g). These images exemplify the wide spectrum of lesion appearances, encompassing variations in color, shape, texture, and other important features.



(a) (b) (c) (d)

(e) (f) (g)

Figure 1. Sample images from the ISIC 2018 dataset; (a) actinic keratosis, (b) basal cell carcinoma, (c) benign keratosis, (d) dermatofibroma, (e) melanoma, (f) melanocytic nevi, and (g) vascular lesions

To ensure the reliability and accuracy of the dataset, each image underwent a rigorous annotation process performed by dermatologists and medical experts. The annotations include the localization and segmentation of lesions, as well as the identification of specific features indicative of different skin diseases. These annotations serve as ground truth labels for the training and evaluation of the models. Moreover, the ISIC 2018 dataset includes additional metadata, such as lesion location, and clinical information. This supplementary information enhances the dataset's richness and allows for potential correlations between various patient characteristics and skin disease outcomes to be explored. The ISIC 2018 dataset poses several challenges, including class imbalance, as certain skin disease categories may have a higher prevalence compared to others. Additionally, the presence of subtle variations in lesion appearance, diverse skin types, and imaging conditions further adds to the complexity of the dataset. However, addressing these challenges provides an opportunity to develop robust AI models capable of accurately detecting and classifying different skin diseases.

3.2. Preprocessing and feature extraction

Prior to model training, we applied preprocessing techniques to standardize the dataset and ensure consistency across the images [23]. This involved resizing the images to a consistent resolution and normalizing pixel values. Additionally, we performed feature extraction to capture relevant visual characteristics and extract informative features from the images [24]. These features served as inputs to the AI model during training and prediction.

3.3. Transfer learning

In this subsection, we delve into the use of the EfficientNetV2 models as the foundation for our transfer learning approach. The EfficientNetV2 models belong to a family of convolutional neural networks (CNNs), that have been specifically designed to enhance parameter efficiency and training speed. These models are built upon the successes of the EfficientNetV1 models, incorporating a scaling method and a neural architecture search algorithm. The aim of this approach is to jointly optimize the network's architecture and hyperparameters. In Figure 2, we illustrate the fundamental operations used to construct the EfficientNetV2 models, namely the MBConv and fused-MBConv structures. These building blocks play a crucial role in the efficient and effective representation learning capabilities of the models.

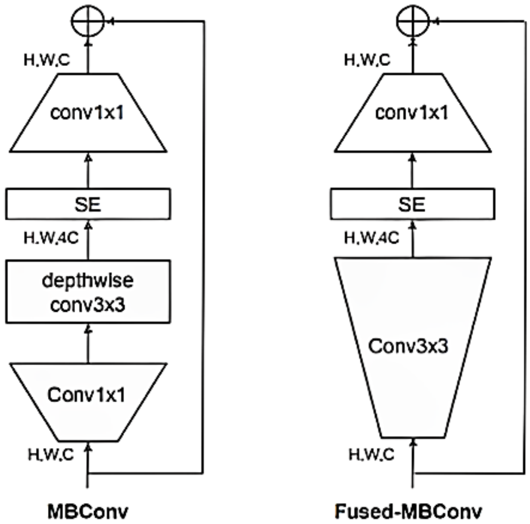


Figure 2. Building blocks of EfficientNetV2 models: MBConv and fused-MBConv structure

EfficientNetV2 models have demonstrated remarkable performance across a wide range of vision tasks, including image classification, object detection, and semantic segmentation. Through their refined architecture and carefully tuned hyperparameters, these models have achieved state-of-the-art results. One notable advantage of EfficientNetV2 models is their computational efficiency. They have been engineered to deliver impressive performance while minimizing the computational resources required for training and inference. This characteristic makes them particularly suitable for applications with limited computational capacity or strict latency requirements [25].

3.4. Model training and evaluation

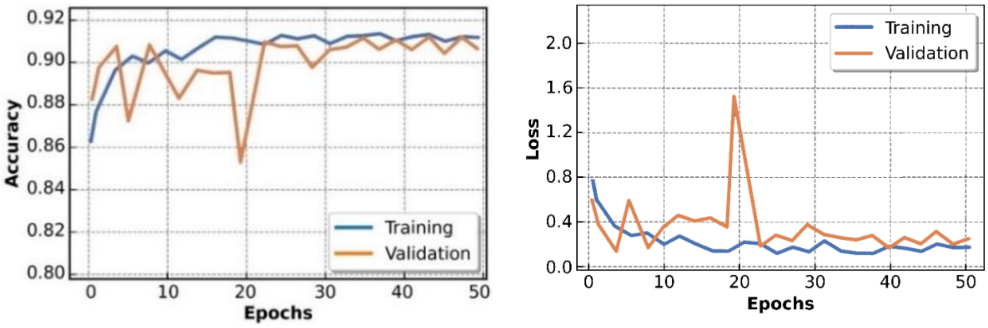
The AI model was trained using a combination of the preprocessed images and extracted features [26], along with the assigned class weights. We employed appropriate training strategies, such as mini-batch stochastic gradient descent, to optimize the model's parameters [27]. To prevent overfitting, we employed techniques such as dropout regularization and early stopping. To assess the performance of the trained model, we conducted rigorous evaluation using various performance evaluation metrics, including accuracy, precision, recall, and F1-score [28]. These metrics provided insights into the model's predictive capabilities and its ability to accurately identify and classify skin diseases. In the next section, we present the experimental results and performance evaluation, shedding light on the impact of data balancing through classes weighting on the transfer learning performance for skin disease prediction.

4. EXPERIMENTAL FINDINGS AND ANALYSIS

In this section, we present the experimental findings and performance evaluation of our proposed approach, which focuses on utilizing classes weighting to improve transfer learning performance for skin disease prediction. We conducted two experiments to assess the effectiveness of our methodology: the baseline model with the default dataset and the TL model with classes weighting.

4.1. Experiment 1: baseline model with the imbalanced dataset

In the first experiment, we trained a baseline model using the default dataset structure without applying any data balancing techniques. This experiment allowed us to establish a performance baseline for comparison against the TL model with classes weighting. The performance evaluation of the baseline pre-trained model on the ISEC 2018 dataset demonstrates notable results. The model underwent training for 25 epochs with consistent configurations to ensure reliable comparisons. The obtained results illustrate the model's capacity to effectively classify the images within the dataset with a high degree of accuracy. To ascertain that the model does not suffer from issues of overfitting or underfitting, we conducted an analysis of the learning and validation curves for the baseline pre-trained model. Figure 3 showcases the accuracy and loss graphs, which provide insight into the model's performance. The training in Figure 3(a) and validation curves in Figure 3(b) are plotted, offering a visual representation of the model's learning progress and generalization capabilities.



(a) (b)

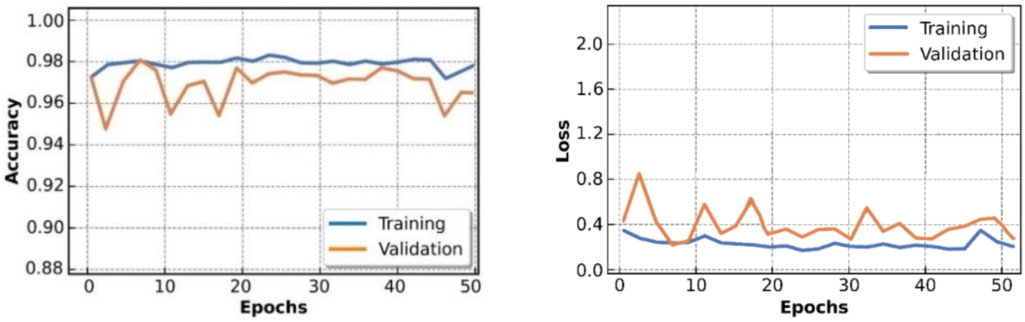
Figure 3. Training progress; (a) accuracy and (b) loss curves for the baseline model

The accuracy graph depicts the upward trend of the model's accuracy on both the training and validation sets, indicating its ability to correctly classify the images, with the maximum accuracy being roughly 91-92%. Meanwhile, the loss graph demonstrates a declining trend, signifying the gradual reduction of the model's training loss and validation loss, with the least amount of loss reaching roughly 0.25-0.3. These observations further validate the robustness and effectiveness of the baseline pre-trained model on the imbalanced ISEC 2018 dataset. Despite the challenge of imbalanced data, the confusion matrix obtained by this model demonstrates significant values for the TP, TN, FP, and FN categories. These values indicate that the model has effectively captured the presence or absence of skin diseases, despite the data imbalance issue.

4.2. Experiment 2: TL model with classes weighting

In the second experiment, we trained the TL model using the same dataset but with the inclusion of classes weighting. This approach aimed to address the class imbalance challenge and improve the overall predictive performance of the model. To implement classes weighting, we calculated the weights for each class using the mathematical equations described in section 2. These weights were then incorporated into the loss function during training, assigning higher weights to the minority classes and lower weights to the majority classes. This study aims to delve into the proposed approach and examine how the relationship between the training and validation curves can be utilized to attain optimal performance in addressing the issue of data imbalance. The investigation focuses on harnessing the insights gained from the training and validation curves to effectively resolve the data imbalance problem. To illustrate this concept, Figure 4 presents the accuracy and loss graphs of the pre-trained model employing classes weighting data. The training in Figure 4(a) and validation curves in Figure 4(b) are plotted, allowing for a comprehensive visualization of the model's learning progress and generalization capabilities using the classes weighting approach.

To provide a comprehensive evaluation of the performance of the pre-trained classifier in the multi-class classification task utilizing the balanced dataset, the results were presented in the form of a confusion matrix, as depicted in Figure 5. The confusion matrix serves to clarify and elucidate the classification outcomes by displaying the distribution of predicted classes against the actual classes. This matrix enables a detailed analysis of the classifier's performance in accurately classifying various types of skin diseases. By examining the confusion matrix, valuable insights can be gained regarding the effectiveness and efficacy of the pre-trained classifier for skin disease classification.



(a) (b)

Figure 4. Training and validation curves for the pre-trained model with classes weighting: (a) accuracy and (b) loss curves

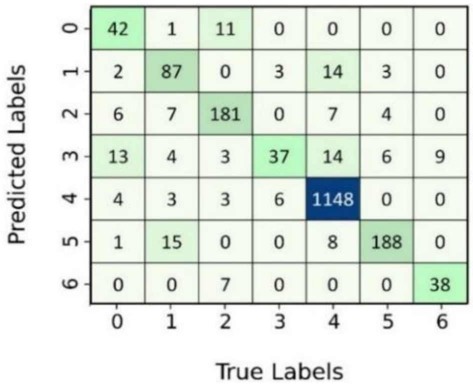


Figure 5. Confusion matrix of the pre-trained model with balanced dataset

To further elucidate these findings, Table 2 encompasses the evaluation metrics employed to gauge the performance of the models. These metrics serve as quantitative measures to assess different aspects of the models' performance, including accuracy, precision, recall, and F1-score. By incorporating these evaluation metrics into Table 2, a comprehensive summary of the models' performance is presented, facilitating a comparative analysis of their efficacy in predicting skin diseases.

Table 2. Quantitative performance assessment of the models in skin disease prediction

Models	Loss (%)	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-score (%)
Experiment 1 (Imbalanced data)	8.57	92.27	51.67	94.65	47.54	66.84
Experiment 2 (Balanced data)	3.01	98.04	84.27	99.01	87.36	91.04

4.3. Results and analysis

The results of the experiments demonstrated the significant impact of utilizing classes weighting in improving the transfer learning performance for skin disease prediction. In the baseline model, trained without data balancing, we observed lower accuracy, precision, recall, and F1-score compared to the TL model with classes weighting. This outcome indicated that the default dataset structure alone was insufficient in effectively handling the class imbalance present in skin disease datasets. The model struggled to accurately predict the minority classes, leading to imbalanced, and biased results. In contrast, the TL model with classes weighting exhibited notable improvements across all evaluation metrics. By incorporating classes weighting, the model demonstrated enhanced sensitivity in identifying positive instances from the minority classes, resulting in a more reliable and accurate prediction of skin diseases.

4.4. Discussion

The experimental results highlighted the significance of data balancing through classes weighting in enhancing the predictive capabilities of transfer learning models for skin disease prediction. By appropriately adjusting the weights assigned to different classes, we achieved more balanced and accurate predictions across all skin disease categories. The findings of this study have practical implications for the field of dermatology and healthcare. Accurate and early detection of skin diseases is crucial for timely interventions and effective treatment planning. Our proposed approach, with its improved transfer learning performance and class-balanced predictions, can serve as a valuable tool to support dermatologists and healthcare professionals in accurate diagnosis and decision-making processes.

5. CONCLUSION AND FUTURE DIRECTIONS

In this study, we have investigated the efficacy of data balancing through classes weighting to enhance TL performance for skin disease prediction. The results obtained from our experiments and performance evaluation have provided valuable insights and significant outcomes, underscoring the effectiveness of our proposed approach. Through our experiments, it was observed that data balancing through classes weighting plays a crucial role in mitigating the adverse effects of class imbalance on transfer learning models. By appropriately assigning weights to different classes, we achieved more balanced predictions across various skin disease categories.

Looking ahead, there are several promising avenues for future research in the fields of data balancing and transfer learning for skin disease prediction. Exploring additional data balancing techniques, such as oversampling or undersampling, could offer further insights and performance improvements. Furthermore, the integration of multi-modal data and evaluation of external validation datasets could provide a more comprehensive understanding of skin diseases and validate the effectiveness of our proposed approach.

ACKNOWLEDGEMENTS

This research project was independently financed by the authors, who bore all financial responsibilities. The authors were fully committed to providing the necessary resources and support throughout the study.

REFERENCES

[1] B. Shetty, R. Fernandes, A. P. Rodrigues, R. Chengoden, S. Bhattacharya, and K. Lakshmanna, "Skin lesion classification of dermoscopic images using machine learning and convolutional neural network," *Scientific Reports*, vol. 12, no. 1, pp. 1–11, Oct. 2022, doi: 10.1038/s41598-022-22644-9.

[2] M. A. Al-masni, D.-H. Kim, and T.-S. Kim, "Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification," *Computer Methods and Programs in Biomedicine*, vol. 190, Jul. 2020, doi: 10.1016/j.cmpb.2020.105351.

[3] B. Krawczyk, "Learning from imbalanced data: open challenges and future directions," *Progress in Artificial Intelligence*, vol. 5, no. 4, pp. 221–232, Nov. 2016, doi: 10.1007/s13748-016-0094-0.

[4] H. A. Khorshidi and U. Aickelin, "Constructing classifiers for imbalanced data using diversity optimisation," *Information Sciences*, vol. 565, pp. 1–16, Jul. 2021, doi: 10.1016/j.ins.2021.02.069.

[5] W. Zhang, X. Li, X.-D. Jia, H. Ma, Z. Luo, and X. Li, "Machinery fault diagnosis with imbalanced data using deep generative adversarial networks," *Measurement*, vol. 152, Feb. 2020, doi: 10.1016/j.measurement.2019.107377.

[6] M.-A. Ouassil, B. Cherradi, S. Hamida, M. Errami, O. E. Gannour, and A. Raihani, "A Fake News Detection System based on Combination of Word Embedded Techniques and Hybrid Deep Learning Model," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 10, 2022, doi: 10.14569/IJACSA.2022.0131061.

[7] P. Vuttipittayamongkol, E. Elyan, and A. Petrovski, "On the class overlap problem in imbalanced data classification," *Knowledge-Based Systems*, vol. 212, Jan. 2021, doi: 10.1016/j.knosys.2020.106631.

[8] Y. Lamalem, S. Hamida, Y. Tazouti, O. El Gannour, K. Housni, and B. Cherradi, "Evaluating multi-state systems reliability with a new improved method," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 3, pp. 1568–1576, Jun. 2022, doi: 10.11591/eei.v11i3.3509.

[9] H. Liu, M. Zhou, and Q. Liu, "An embedded feature selection method for imbalanced data classification," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 3, pp. 703–715, May 2019, doi: 10.1109/JAS.2019.1911447.

[10] H. M. Balaha and A. E.-S. Hassan, "Skin cancer diagnosis based on deep transfer learning and sparrow search algorithm," *Neural Computing and Applications*, vol. 35, no. 1, pp. 815–853, Jan. 2023, doi: 10.1007/s00521-022-07762-9.

[11] S. Fotouhi, S. Asadi, and M. W. Kattan, "A comprehensive data level analysis for cancer diagnosis on imbalanced data," *Journal of Biomedical Informatics*, vol. 90, Feb. 2019, doi: 10.1016/j.jbi.2018.12.003.

[12] W. Zhang, X. Li, H. Ma, Z. Luo, and X. Li, "Universal Domain Adaptation in Fault Diagnostics With Hybrid Weighted Deep Adversarial Learning," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 12, pp. 7957–7967, Dec. 2021, doi: 10.1109/TII.2021.3064377.

[13] G. Douzas, F. Bacao, and F. Last, "Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE," *Information Sciences*, vol. 465, pp. 1–20, Oct. 2018, doi: 10.1016/j.ins.2018.06.056.

[14] Z. Mungloo-Dilmohamud, M. Heenaye-Mamode Khan, K. Jhumka, B. N. Beedassy, N. Z. Mungloo, and C. Peña-Reyes, "Balancing Data through Data Augmentation Improves the Generality of Transfer Learning for Diabetic Retinopathy Classification," *Applied Sciences*, vol. 12, no. 11, pp. 1–17, May 2022, doi: 10.3390/app12115363.

[15] A. Luque, A. Carrasco, A. Martín, and A. de las Heras, "The impact of class imbalance in classification performance metrics based on the binary confusion matrix," *Pattern Recognition*, vol. 91, pp. 216–231, Jul. 2019, doi: 10.1016/j.patcog.2019.02.023.

[16] T. Diwan, R. Shukla, E. Ghuse, and J. V. Tembhurne, “Model hybridization & learning rate annealing for skin cancer detection,” *Multimedia Tools and Applications*, vol. 82, no. 2, pp. 2369–2392, Jan. 2023, doi: 10.1007/s11042-022-12633-5.

[17] P. N. Srinivasu, J. G. SivaSai, M. F. Ijaz, A. K. Bhoi, W. Kim, and J. J. Kang, “Classification of Skin Disease Using Deep Learning Neural Networks with MobileNet V2 and LSTM,” *Sensors*, vol. 21, no. 8, pp. 1–27, Apr. 2021, doi: 10.3390/s21082852.

[18] M. A. Mahjoubi, S. Hamida, O. El Gannour, B. Cherradi, A. El Abbassi, and A. Raihani, “Improved Multiclass Brain Tumor Detection using Convolutional Neural Networks and Magnetic Resonance Imaging,” *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 3, pp. 406–414, 2023, doi: 10.14569/IJACSA.2023.0140346.

[19] H. Dalianis, “Evaluation Metrics and Evaluation,” in *Clinical Text Mining*, Cham: Springer International Publishing, 2018, pp. 45–53, doi: 10.1007/978-3-319-78503-5_6.

[20] L. Abdi and S. Hashemi, “To Combat Multi-Class Imbalanced Problems by Means of Over-Sampling Techniques,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 1, pp. 238–251, Jan. 2016, doi: 10.1109/TKDE.2015.2458858.

[21] T. M. Alam *et al.*, “An Efficient Deep Learning-Based Skin Cancer Classifier for an Imbalanced Dataset,” *Diagnostics*, vol. 12, no. 9, pp. 1–16, Aug. 2022, doi: 10.3390/diagnostics12092115.

[22] P. Tschandl, C. Rosendahl, and H. Kittler, “The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions,” *Scientific Data*, vol. 5, no. 1, pp. 1–9, Aug. 2018, doi: 10.1038/sdata.2018.161.

[23] N. M. Nawi, W. H. Atomi, and M. Z. Rehman, “The Effect of Data Pre-processing on Optimized Training of Artificial Neural Networks,” *Procedia Technology*, vol. 11, pp. 32–39, 2013, doi: 10.1016/j.protcy.2013.12.159.

[24] A. Smolinska, A.-C. Hauschild, R. R. R. Fijten, J. W. Dallinga, J. Baumbach, and F. J. van Schooten, “Current breathomics—a review on data pre-processing techniques and machine learning in metabolomics breath analysis,” *Journal of Breath Research*, vol. 8, no. 2, pp. 39–41, Apr. 2014, doi: 10.1088/1752-7155/8/2/027105.

[25] S. Saleh, B. Cherradi, O. El Gannour, N. Gouiza, and O. Bouattane, “Healthcare monitoring system for automatic database management using mobile application in IoT environment,” *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 2, pp. 1055–1068, Apr. 2023, doi: 10.11591/eei.v12i2.4282.

[26] S. Hamida, B. Cherradi, O. El Gannour, A. Raihani, and H. Ouajji, “Cursive Arabic handwritten word recognition system using majority voting and k-NN for feature descriptor selection,” *Multimedia Tools and Applications*, Mar. 2023, doi: 10.1007/s11042-023-15167-6.

[27] J. Fürnkranz *et al.*, “Model Evaluation,” in *Encyclopedia of Machine Learning*, Boston, MA: Springer US, 2011, pp. 683–683, doi: 10.1007/978-0-387-30164-8_550.

[28] B. J. Biggerstaff, “Confidence intervals for the difference of two proportions estimated from pooled samples,” *Journal of Agricultural, Biological, and Environmental Statistics*, vol. 13, no. 4, pp. 478–496, Dec. 2008, doi: 10.1198/108571108X379055.

BIOGRAPHIES OF AUTHORS

Oussama El Gannour is a 28-year-old Moroccan from Kenitra. In 2018, he obtained a master’s specialized degree in internet of things and mobile services from the National Higher School of Computer Science and Systems Analysis of Rabat, Mohammed V University. In 2019, he is a Ph.D. student at the Research Laboratory on Electrical Engineering and Intelligent Systems of the Hassan II University of Casablanca. His research interests contribute to the integration of the internet of things and artificial intelligence in the medical field. His research works published include screening patients with heart disease, and COVID19 using AI techniques, and IoT. He can be contacted at email: oussama.elgannour@gmail.com.

Soufiane Hamida, a 29-year-old researcher from Rabat, Morocco, is highly knowledgeable in the field of machine learning methodologies for pattern recognition, as evidenced by his Ph.D. degree. He further honed his skills and expertise in the field of educational technology by obtaining his master’s degree from the Higher Normal School of Tetouan, Abdelmalek Essaadi University in 2017. Currently, he is actively contributing to the advancement of research at the Electrical Engineering and Intelligent Systems Research Laboratory at Hassan II University of Casablanca, Morocco. Furthermore, he is making significant efforts towards furthering research at the GENIUS Laboratory at SupMTI in Rabat, Morocco. He has made significant contributions to several research works in the areas of handwritten character and word recognition and optimization of machine learning models. He can be contacted at email: hamida.93s@gmail.com.

Yasser Lamalem was born in Kenitra, Morocco. He received the master’s degree in computer science from Ibn Tofail University, in 2016, where he is currently pursuing the Ph.D. degree in network reliability. His current research interests include multi-state network’s reliability and binarystate network’s reliability. He can be contacted at email: yasserlamalem@gmail.com.

Mohamed Amine Mahjoubi was born in 1996 in Jorf. In 2022, he obtained a master's specialized degree in computer science and complex systems engineering from the Faculty of Science and Technology of Errachidia, Moulay Ismail University. In 2022, he will be a Ph.D. student at the Research Laboratory on Electrical Engineering and Intelligent Systems of the Hassan II University of Casablanca. His research interests contribute to the integration of AI in the medical field. His published research includes screening patients with brain diseases and Alzheimer's disease using deep learning techniques. He can be contacted at email: aminemhj17@gmail.com.

Bouchaib Cherradi was born in 1970 at El Jadida, Morocco. He received the B.S. degree in Electronics in 1990 and the M.S. degree in Applied Electronics in 1994 from the ENSET Institute of Mohammedia, Morocco. He received the DESA diploma in Instrumentation of Measure and Control (IMC) from Chouaib Doukkali University at El Jadida in 2004. He received his Ph.D. in Electronics and Image processing from the Faculty of Science and Technology, Mohammedia. His research interests include applications of Massively parallel architectures, cluster analysis, pattern recognition, image processing, fuzzy logic systems, AI, machine learning and deep learning in medical and educational data analysis. Dr. Cherradi works actually as an associate professor in CRMEF-El Jadida. In addition, he is associate researcher member of Electrical Engineering and Intelligent Systems (EEIS) Laboratory in ENSET Mohammedia, Hassan II University of casablanca (UH2C). He is a supervisor of several Ph.D. students. He can be contacted at email: bouchaib.cherradi@gmail.com.

Abdelhadi Raihani was appointed as a professor in Electronics Engineering at Hassan II University of Casablanca, ENSET Institute, Mohammedia Morocco since 1991. He received the B.S. degree in Electronics in 1987 and the M.S. degree in Applied Electronics in 1991 from the ENSET Institute. He received his DEA diploma in information processing from the Ben M'sik University of Casablanca in 1994. He received the Ph.D. in Parallel Architectures Application and image processing from the Ain Chock University of Casablanca in 1998. His current research interests are in the medical image processing areas, electrical engineering fields, particularly in renewable energy, energy management systems, and power and energy systems control. He is an active member in national research programs with IRESEN under the grant "Green INNO Project/UPISREE". He can be contacted at email: abraihani@yahoo.fr.