Advances in Data Science and Adaptive Analysis
(2025) 2550003 (12 pages)

DOI: 10.1142/S2424922X25500032

Advanced Skin Cancer Classification Using Yet Another Hill Climbing

Yoshiyasu Takefuji

Faculty of Data Science, Musashino University
3-3-3 Ariake Koto-ku, Tokyo 135-8181, Japan
takefuji@keio.jp

Received 6 July 2024
Accepted 6 January 2025
Published 6 February 2025

This paper introduces a novel hill-climbing technique aimed at maximizing the prediction accuracy in skin cancer classification. It highlights the efficacy of epoch–baseline termination management coupled with data augmentation, especially when dealing with the severely imbalanced HAM10000 data set. The method alters the hill-climbing function's termination condition, enabling the system to evade local minima and find superior solutions. Termination conditions using baseline and epochs significantly enhance prediction accuracy, while data augmentation balances the data set. The method achieves a prediction accuracy of over 0.99 in classifying seven types of skin cancers using the HAM10000 data set, as validated by repetitive random cross-validation and the confusion matrix. The paper concludes that the proposed method, when combined with data augmentation, can enhance deep learning and is applicable to oncology classification. The advanced skin classification algorithm proposed herein has achieved the highest prediction accuracy in the benchmark with the HAM10000 data set, which can be used for cancer research classification in general.

Keywords: Severely imbalanced data; skin cancer classification; data augmentation; termination condition.

Highlights

(1) The method excels in benchmark with high accuracy in skin cancer classification.
(2) Combining hill climbing and data augmentation boosts the CNN performance.
(3) Optimal termination conditions enhance the accuracy with data augmentation.
(4) Validating with a confusion matrix provides evidence to support the proposed claims.

1. Introduction

Clinical training is essential for specialists to receive on-the-job training. However, there is a lack of versatility in practicing clinical training alone. With the advent of

artificial intelligence (AI), it has become possible to train practitioners more effectively. This paper proposes an AI trainer with the highest accuracy for skin cancer classification. The precise classification of skin cancer is of utmost importance as it dictates the specific medical treatment required. Consequently, the proposed method can aid users in identifying one among the seven types of skin cancer. The proposed method also holds potential for application in general image classification tasks.

The classification of skin cancer is of paramount importance for a variety of reasons. Early diagnosis through correct classification of skin lesions can aid clinical decision-making by providing an accurate disease diagnosis, potentially increasing the chances of cure before the cancer spreads. The classification of cancer by anatomic disease extent, or stage, is a major determinant of appropriate treatment and prognosis. Risk assessment is another crucial aspect, as knowing your skin type can help you make better decisions about how to protect your skin in the sun, given that multiple sunburns have been linked to the development of skin cancer. In the realm of research and development, high-throughput genomic technologies have facilitated the genomic, transcriptomic, and epigenomic profiling of several cancers, including melanoma. This knowledge can help scientists identify the most suitable technology to address melanoma-related questions. Lastly, genomic technologies might allow for better prediction of the biological and, subsequently, clinical behaviors for each subset of melanoma patients. They may even identify all molecular changes in tumor cell populations during disease evolution, paving the way for personalized medicine.

This paper presents an innovative algorithm that achieves a benchmark accuracy of over 0.99 in skin cancer classification, outperforming a variety of existing heuristic algorithms. This superior performance is demonstrated when the algorithm is tested on one of the most extensive data sets available. However, it's important to note that this data set is imbalanced, a condition where conventional methods often fall short in providing satisfactory solutions. The results of our study highlight the effectiveness of our proposed algorithm in handling such imbalanced data sets. The size of the test set is determined by the size of the smallest class in the data set, meaning that the largest possible test set was used for this validation. This approach ensures the most rigorous testing and validation of our algorithm.

This paper showcases the potency of a novel hill-climbing method that integrates data augmentation. Although data augmentation, a technique capable of transforming severely imbalanced data sets into balanced ones, is not a novel approach, this paper illustrates how the amalgamation of epoch–baseline termination management and data augmentation can bolster machine learning performance. The proposed method holds promise for a broad spectrum of classification tasks within the field of oncology.

Hill climbing is a heuristic search or informed search technique having different weights or variants used for mathematical optimization problems in the field of artificial intelligence. This paper proposes a new hill-climbing technique. Changing

the termination condition forces the system state to escape from the local minima and reach a better solution. The effect of the terminal condition changes is similar to the hill-climbing function. The termination conditions using baseline and epochs significantly play a key role in improving prediction accuracy.

According to the American Cancer Society, estimates for melanoma in the United States for 2021 were as follows: 106,110 new melanomas would be diagnosed about 62,260 in men and 43,850 in women, and 7,180 people were expected to die of melanoma about 4,600 men and 2,580 women [American Cancer Society (2021)].

Tschandl *et al.* [2018] released the HAM10000 ("Human Against Machine with 10,015 training images") data set with seven kinds of skin cancers: melanoma (mel), melanocytic nevus (nv), basal cell carcinoma (bcc), actinic keratosis/Bowen's disease (intraepithelial carcinoma) (akiec), benign keratosis (solar lentigo/seborrheic keratosis/lichen_planus-like_keratosis) (bkl), dermatofibroma (df), and vascular lesion (vasc) as defined by the International Dermatology Society. There were duplications and triplications in the HAM10000 data set. However, the same HAM numbers were used to identify them while the ISIC codes are all unique.

Table 1 shows the numbers of images of seven cancers with imbalanced data. Imbalanced data typically refers to a classification problem where the number of observations per class is not equally distributed. Data are said to suffer the class imbalance problem when the class distributions are highly imbalanced. The HAM10000 is a severely imbalanced data set. The conventional cross-validation cannot be used in a severely imbalanced data set where the largest imbalance ratio is 6705 (nv)/115 (df) = 58. A strong imbalance ratio can deteriorate the quality of a classifier in general. In this paper, repetitive random validation is used for maximizing the prediction accuracy of the proposed algorithm using the severely imbalanced data set.

Many of the existing methods attempted to improve individual algorithms using deep learning frameworks by changing the architectures and filters without data augmentation.

Data augmentation is a technique used to increase the amount of data by adding slightly modified copies of existing data or newly created synthetic data from the existing data. Data augmentation is particularly useful for enriching minority classes

Table 1. HAM10000 data set and the numbers of images.

Cancer type	Number of images
nv	6,705
mel	1,113
bkl	1,099
bcc	514
akiec	327
vasc	142
df	115

by generating additional instances of these classes. Note that after eliminating test data from the HAM10000 data set, data augmentation is only applied to training data.

This paper shows the effectiveness of data augmentation for the classification of seven skin cancers using the HAM10000 data set. A comprehensive review of the literature was carried out to evaluate the accuracy achieved in skin cancer classification using the identical data set. This review provides a comparative analysis of the performance metrics across various studies.

Sai Charan *et al.* [2020] proposed a skin lesion classification model with convolutional neural networks (CNNs) using preprocessing techniques including brightening the images using piece-wise linear transformation function, grayscale conversion of the image, and resizing the image, respectively, for analyzing the dermoscopic images. The model, enhanced by data augmentation, segmentation, and 10-fold cross-validation, also employed preprocessing techniques like image brightening, grayscale conversion, and resizing. The model's best accuracy was 0.886.

Lucius *et al.* [2020] reported eight deep neural frameworks where the highest accuracy for seven-skin-cancer classification was 0.8247 using a MobileNet model. Their model was trained on 8,015 images from the ISIC Archive and tested on 2,003 images, improving the accuracy when clinical data was added. Their model outperformed general practitioners in classifying seven pigmented skin lesions, both at low and high image resolutions. Their study aimed to incorporate these AI tools into general practice to enhance diagnostic accuracy, particularly where the high-resolution equipment was unavailable [Lucius *et al.* (2020)].

Le *et al.* [2020] showed a modified ResNet50 deep learning model with the highest accuracy of 0.94, while the outperformed dermatologists had an accuracy of 0.84. An end-to-end deep learning process was used, employing transfer learning and multiple pre-trained models. The classification process was enhanced with class-weighted and focal loss. Their result was an ensemble of modified ResNet50 models that could classify skin lesions into one of the seven classes with top-1, top-2, and top-3 accuracies of 93%, 97%, and 99%, respectively. Their system had the potential to be integrated into computer-aided diagnosis systems, aiding dermatologists in skin cancer diagnosis. However, their results did not demonstrate a high level of accuracy for the less-prevalent classes of skin cancer [Le *et al.* (2020)]. It must be kept in mind that their proposed method boasted the highest accuracy across all categories of skin cancer.

Thakar and Kulkarni [2020] proposed a CNN method with an accuracy of 0.78. Their machine learning strategies employed image processing techniques to identify moles in a given image. Their CNN-based approach was for image classification, achieving an optimal overall accuracy of 78.08% and a commendable multiclass AUC for all classes in the HAM10000 data set.

Nasiri *et al.* [2020] reported a deep convolutional neural networks approach with an accuracy of 0.75. Their approach used the DePicT Melanoma Deep-CLASS system to retrieve new input images from its case base, providing users with

more accurate recommendations related to their specific problem (e.g. image of the affected area). The system's efficiency was confirmed through the use of the ISIC Archive data set for the classification of skin lesions as benign and malignant melanoma. The core of DePicT Melanoma Deep-CLASS was a CNN with 16 layers (excluding the input and output layers), which could be trained and learnt recursively [Nasiri *et al.* (2020)].

Hassan *et al.* [2020]showed that an accuracy of 0.92 was achieved by densely connected convolutional network (DenseNet-121). They utilized preprocessing and augmentation methods, along with the implementation of DenseNet-121 for the training processes. Their training set was trained on a pre-trained DenseNet-121 model. In their sequential model, they incorporated GlobalAveragePooling2D, Dropout, and a dense layer with a "softmax" activation function. The Dropout method was a simple yet effective way to prevent overfitting in the neural network. Their model consisted of a total of 7,044,679 parameters, out of which 6,961,031 parameters were trainable and 83,648 parameters were nontrainable [Hassan *et al.* (2020)].

Ratul *et al.* [2020] exhibited the highest classification accuracy of 0.8981 using dilated convolution inception V3 model. In order to implement dilated convolution, they opted for transfer learning using four renowned architectures: VGG16, VGG19, MobileNet, and InceptionV3. The top-1 accuracies attained on the dilated versions of VGG16, VGG19, MobileNet, and InceptionV3 were 87.42%, 85.02%, 88.22%, and 89.81%, respectively. Dilated InceptionV3 demonstrated the highest classification accuracy, recall, precision, and F_1-score, while dilated MobileNet also showed high classification accuracy with the minimal computational complexities. Dilated InceptionV3 outperformed all known methods in terms of the overall and per-class accuracies in skin lesion classification, even when tested on a complex open-source data set with class imbalances.

The review results suggest that our proposed method significantly outperforms the existing methods using the identical data set.

This paper's key contribution lies in the scrutiny of highly imbalanced data sets, with a specific focus on the smallest class among seven distinct classes. The proposed hill-climbing method's efficacy is substantiated through the application of cross-validation methodologies on a randomly chosen test data set. In other words, the smallest class comprises 115 instances. Of these, 87 instances are utilized for machine learning. The remaining 28 instances, derived from subtracting 87 from 115, along with an additional 168 instances (calculated as 28 instances from each of the other six classes), make up a total of 196 instances. These are used for test validation to avoid data leakage. This selection process is carried out randomly and repetitively to ensure a robust validation procedure.

2. Methods

This paper shows the significant difference between CNN with data augmentation and that without data augmentation for skin cancer classification.

The proposed program using HAM10000 data set is based on the Keras framework using the following CNN model:

```
model = Sequential()
model.add(Conv2D(32, (3, 3), activation='relu', input_shape=(64,64,3)))
model.add(MaxPool2D((2, 2)))
model.add(Conv2D(64, (3, 3), activation='relu'))
model.add(MaxPool2D((2, 2)))
model.add(Conv2D(128, (3, 3), activation='relu'))
model.add(MaxPool2D((2, 2)))
model.add(Flatten())
model.add(Dense(1000, activation='relu'))
model.add(Dense(7))
model.add(Activation(tf.nn.softmax))
```

2.1. *How to handle an imbalanced data set*

It is crucial to handle the training and testing of a severely imbalanced data set properly. Assume an extreme example with seven types in a data set with a total of 10,000 images where each of the six types has 10 images, respectively, and the seventh type has 9,940 images. When splitting the data set into a testing data set and a training data set randomly in proportion to the number of respective images, the training and testing data sets will consist mostly of the seventh type. Any algorithm can have a very high prediction accuracy for the seventh classification and a very low prediction accuracy for the other six classifications. In other words, if there is a strong bias in the number of training data per class or imbalanced data, the neural network will learn in a biased manner. Therefore, in severely imbalanced data sets, special care should be taken when splitting the training data and testing data.

We must understand that the more the images that are used for training, the better the training accuracy that can be achieved. In a severely imbalanced data set, the smaller the training data set being used, the less successful machine learning will be. Before converting a severely imbalance data set into a balanced one, splitting the imbalanced data set into testing data and training data, respectively, is needed. Therefore, data augmentation is only applied to the training data. Remember that the test set size is determined by the smallest class (dermatofibroma) in the skin cancer classes.

By converting a severely imbalanced training data into a balanced one, the convolutional neural network performs unbiased learning.

A confusion matrix is a table that is often used to describe the performance of a classification model on a set of test data for which the true values are known. In

the confusion matrix, precision, recall, and F_1-score are calculated for each of the seven skin cancer classes, respectively.

In the classification, there are four types: true positive (tp), true negative (tn), false positive (fp), and false negative (fn).

A true positive is an outcome where the model correctly predicts the positive class. A true negative is an outcome where the model correctly predicts the negative class. A false positive is an outcome where the model incorrectly predicts the positive class. And a false negative is an outcome where the model incorrectly predicts the negative class.

Precision is defined as the number of true positives divided by the number of true positives plus the number of false positives,

$$\text{Precision} = \text{tp}/(\text{tp} + \text{fp}).$$

Recall or sensitivity is defined as the number of true positives divided by the number of true positives plus the number of false negatives,

$$\text{Recall} = \text{tp}/(\text{tp} + \text{fn}).$$

The F_1-score can be interpreted as a weighted average of precision and recall. The F_1-score is defined by

$$F_1 - \text{score} = 2(\text{Precision})(\text{Recall})/(\text{Precision} + \text{Recall}) = 2\text{tp}/(2\text{tp} + \text{fp} + \text{fn}).$$

Finally, accuracy is defined by

$$\text{Accuracy} = (\text{tp} + \text{tn})/(\text{tp} + \text{tn} + \text{fp} + \text{fn}).$$

2.2. *Rigorous testing and strategy*

For severely imbalanced data sets, the traditional cross-validation is not applicable. When the training set is small and severely imbalanced, it usually ignores the overall class distribution, leading to information loss. Imbalance is a very common problem in today's scenario causing serious deviations in the performance of traditional neural network classifiers.

Remember that the largest imbalance ratio is 6705 (nv)/115 (df) = 58, while the number of df data is only 115. Therefore, in this paper, we picked $28(= 4 \times 7)$ test images randomly selected from 10,015 images, where four images from every skin cancer of the seven classes (akiec, bcc, bkl, df, nv, vasc, and mel) were randomly chosen. Since there are no duplicate images, there is no leakage of data between the training and test sets.

We know an important fact that the more the data that is trained, the better the classification prediction accuracy that is statistically obtained. For severely imbalanced data sets, we should pay attention to the smallest data class for avoiding a biased learning.

Because dermatofibroma has only 115 images in HAM10000 which is the smallest data of the seven cancers, randomly selecting four images from the 115 images would produce about 3.5% of the test data. In severely imbalanced data, we must

maximize and improve the classification prediction accuracy from the viewpoint of the smallest data class (df). The same number of testing images per class should be used for rigorous testing.

keras_skin64S.py [GitHub (2024)] is a skin cancer classification program with data-augmented training images and 28 testing images selected, where four images are randomly selected from each of the seven skin cancer classes.

Figure 1 shows how 10,015 images using HAM10000 skin cancer data set can be split into an augmented training data set and a 28-image testing data set, respectively.

Remember that after splitting a severely imbalanced data set into testing data and training data, data augmentation is only applied to the training data. The balanced training data is used for machine learning using the proposed CNN.

To verify the robustness of the proposed algorithm, the keras_skin64S.py program was iterated 10 times for validation testing. In other words, the total of $280(= 28 \times 10)$ images were randomly selected and tested, where in each test 28 images were examined as to whether they are correctly classified into seven skin cancers.

In order to run and verify keras_skin64S.py, the following library versions must be met in order to confirm the proposed result:

Keras: 2.2.4

Keras-Applications: 1.0.8

Keras-Preprocessing: 1.1.0

TensorFlow: 1.14.0

TensorFlow-Estimator: 1.14.0

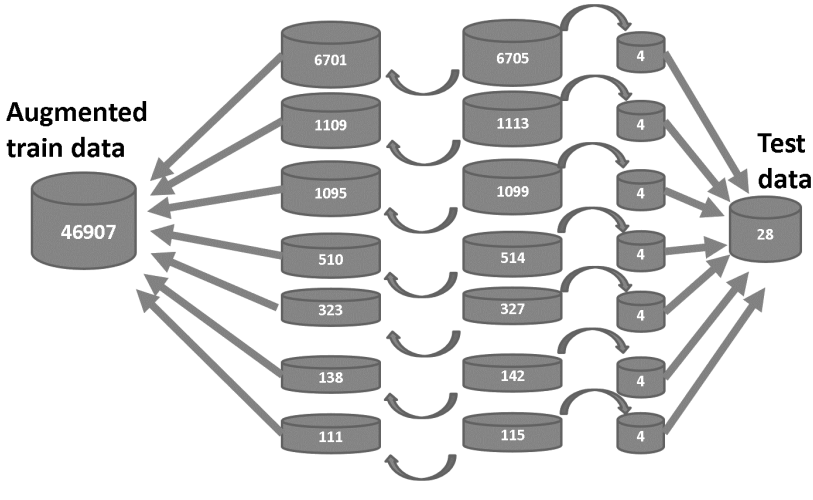

Fig. 1. Rigorous test data and augmented training data using the HAM10000 skin cancer data set.

TensorFlow-GPU: 1.14.0

NumPy: 1.19.1

MKL-Random: 1.1.1

3. Results and Discussion

Accuracy of keras_skin64RGB.py [GitHub (2024)] without data augmentation and that of keras_skin64RGBs.py [GitHub (2024)] with data augmentation using the Synthetic Minority Over-sampling Technique (SMOTE) are shown in Tables 2 and 3, respectively. SMOTE library is used for data augmentation in this paper. The program of keras_skin64RGB.py uses 64 (pixels) $\times$ 64 (pixels) $\times$ 3 (RGB) data features per image for machine learning.

The confusion matrix in Table 2 shows that nv cancer with the largest samples has the highest values for precision, recall, and F_1-score, respectively. This means that nv cancer is well trained, while the other six cancers are poorly trained because of imbalanced data set.

The confusion matrix in Table 3 shows that the augmented data set gives better results than that using the imbalanced data set in Table 2 from the viewpoint of precision, recall, F_1-score, and accuracy, respectively.

Table 2. Confusion matrix generated by keras_skin64RGB.py.

	Precision	Recall	F_1-score	Support
akiec	0.42	0.34	0.38	71
bcc	0.42	0.71	0.53	89
bkl	0.53	0.39	0.45	224
df	0.60	0.38	0.46	24
nv	0.85	0.91	0.88	1,353
vasc	0.81	0.45	0.58	29
mel	0.41	0.27	0.32	213
Accuracy	0.74	2,003		
Macro avg	0.58	0.49	0.51	2,003
Weighted avg	0.73	0.74	0.73	2,003

Table 3. Confusion matrix generated by keras_skin64RGBs.py.

	Precision	Recall	F_1-score	Support
akiec	0.99	1	1	1,354
bcc	0.97	1	0.98	1,349
bkl	0.97	0.95	0.96	1,308
df	1	1	1	1,355
nv	0.91	0.91	0.91	1,347
vasc	1	1	1	1,336
mel	0.95	0.93	0.94	1,338
Accuracy	0.97	9,387		
Macro avg	0.97	0.97	0.97	9,387
Weighted avg	0.97	0.97	0.97	9,387

Regardless of whether one has an imbalanced or balanced data set, we need to pay special attention to rigorous test data.

3.1. *Strategy of hill climbing changing termination conditions*

The proposed hill climbing is based on baseline and epochs. Baseline is a type of model that serves as a benchmark for the capabilities possible with the available data. In other words, the baseline model is essentially a simple model that is used as a reference in machine learning. The number of epochs is a hyperparameter that defines the number of times the learning algorithm processes the entire training data set. In other words, one epoch means that each sample in the training data set has had an opportunity to update its internal model parameters.

In order to improve prediction accuracy, the "baseline" and "epochs" termination conditions play an important role in the classification problem. "Baseline" is a real-number parameter between 0 and 1 that can terminate a running program based on validation accuracy, and "epochs" is a positive integer parameter or the maximum number of epochs.

In the first experiment, the following termination conditions are used: epochs = 400 and baseline = 0.97. The program ends when the "baseline" reaches 0.97 or "epochs" reaches 400.

The classification error for one image yields a prediction classification accuracy of $27/28 = 0.9643$.

Among the 10 iterations in the first experiment, a prediction accuracy of 1 is attained six times, a prediction accuracy of 0.9643 is attained four times, and the average classification prediction accuracy becomes 0.9857.

In the second experiment, we used the termination conditions with epochs = 600 and baseline = 0.97, but if the number of epochs > 300, then we reduced and set the baseline to 0.96 instead of 0.97. In other words, if the number of epochs is up to 300, the validation accuracy is set to 0.97, but if the number of epochs exceeds 300, the validation accuracy is set to 0.96.

Table 4. Confusion matrix generated by keras_skin64S.py.

Epoch 417: Reached baseline, terminating training Balanced = 0.9642857142857143 Precision Recall F_1-score Support				
akiec	1	1	1	4
bcc	1	1	1	4
bkl	1	0.75	0.86	4
df	1	1	1	4
nv	1	1	1	4
vasc	1	1	1	4
mel	0.80	1	0.89	4
Accuracy	0.96	28		
Macro avg	0.97	0.96	0.96	28
Weighted avg	0.97	0.96	0.96	28

Among the 10 iterations in the second experiment, the prediction accuracy of 1 is attained nine times, the prediction accuracy of 0.9643 is attained only once, and the average classification prediction accuracy becomes 0.9964.

Table 4 shows the one-classification error case in the second experiment.

After 10 more iterations of testing keras_skin64S.py in the second experiment, the prediction accuracy was 0.9929 with two misclassifications out of 280 image classifications.

4. Conclusion

This paper demonstrated the effectiveness of the proposed hill climbing with changing the termination conditions of "baseline" for validation accuracy and "epochs" for the maximum number of epochs. The highest classification prediction accuracy over the existing methods was achieved by our proposed program: accuracy=0.9964 when four images were randomly selected from each of the seven cancers and the program was 10 times iterated. In two experiments, out of the 280 images, 279 images or 278 images were successfully classified into seven skin cancers by the proposed method. The proposed system can serve as a tutor for skin cancer classification. The proposed hill climbing can be used for deep learning in general. The proposed Python programs can be made available upon request.

The primary contribution of this paper was the detailed examination of highly imbalanced data sets, particularly focusing on the smallest class among the seven unique classes. The effectiveness of the proposed hill-climbing method was validated through the use of cross-validation techniques on a test data set selected at random. To elaborate, the smallest class consisted of 115 instances. From these, 87 instances were employed for machine learning. The remaining 28 instances, obtained by subtracting 87 from 115, together with an extra 168 instances (calculated as 28 instances from each of the remaining six classes) constituted a total of 196 instances. These instances were used for test validation to prevent data leakage. This selection procedure was performed randomly and repeatedly to ensure a robust validation process.

5. Summary

5.1. *What was already known on the topic*

- Converting an imbalanced data set into a balanced data set can significantly improve the prediction accuracy.
- By data augmentation, convolutional neural networks can learn in an unbiased manner.

5.2. *What this study added to our knowledge*

- Hill climbing, such as the termination conditions using baseline and epochs, plays a key role in significantly improving the prediction accuracy.

- The proposed method achieved the highest prediction accuracy of more than 0.99 for the benchmark classification of seven skin cancers using HAM10000.
- Confusion matrix with cross-validation was used for verifying the proposed claims.

Acknowledgments

The author has been working for the Department of Data Science of Musashino University since April 2021. The author has no conflict of interest with regard to this work. Also, this research received no funding.

ORCID

Yoshiyasu Takefuji https://orcid.org/0000-0002-1826-742X

References

American Cancer Society (2021). Key statistics for melanoma skin cancer. Available at: https://www.cancer.org/cancer/melanoma-skin-cancer/about/key-statistics.html. Accessed on.

GitHub (2024). Python programs. Available at: https://github.com/y-takefuji/ machine-learning-in-medicine/tree/main. Accessed on.

Hassan, S. R., Afroge, S. and Mizan, M. B. (2020). Skin Lesion Classification Using Densely Connected Convolutional Network. *Proc.* 2020 *IEEE Region* 10 *Symp.* (*TENSYMP*), IEEE, Washington, DC, pp. 750–753. doi:10.1109/ TENSYMP50017.2020.9231041.

Le, D. N. T., Le, H. X., Ngo, L. T. and Ngo, H. T. (2020). Transfer learning with class-weighted and focal loss function for automatic skin cancer classification. Preprint, arXiv:2009.05977 [cs.AI].

Lucius, M., De All, J., De All, J. A., Belvisi, M., Radizza, L., Lafranconi, M., Lorenzatti, V. and Galmarini, C. M. (2020). Deep neural frameworks improve the accuracy of general practitioners in the classification of pigmented skin lesions. medRxiv. doi:10.1101/2020.05.03.20072454.

Nasiri, S., Helsper, J., Jung, M. and Fathi, M. (2020). DePicT Melanoma Deep-CLASS: A deep convolutional neural networks approach to classify skin lesion images. *BMC Bioinformatics*, **21**: 84. doi:10.1186/s12859-020-3351-y.

Ratul, Md. A. R., Mozaffari, M. H., Lee, W.-S. and Parimbelli, E. (2020). Skin lesions classification using deep learning based on dilated convolution. bioRxiv. doi:10.1101/860700v3.full.

Sai Charan, D., Nadipinen, H., Sahayam, S. and Jayaraman, U. (2020). Method to classify skin lesions using dermoscopic images. Preprint, arXiv:2008.09418 [eess.IV].

Thakar, P. and Kulkarni, S. A. (2020). Skin lesion classification: A CNN way. *Int. J. Eng. Adv. Technol.*, **9**: 274–278.

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