This full text paper was peer-reviewed at the direction of IEEE Instrumentation and Measurement Society prior to the acceptance and publication.

Skin Cancer Classification based on Cosine Cyclical Learning Rate with Deep Learning

Yali Nie
dept. Electronics Design
Mid Sweden University
Sundsvall, Sweden
yali.nie@miun.se

Marco Carratù
dept. of Industrial Engineering
University of Salerno
Salerno, Italy
mcarratu@unisa.it

Mattias O'Nils
dept. Electronics Design
Mid Sweden University
Sundsvall, Sweden
mattias.onils@miun.se

Paolo Sommella
dept. of Industrial Engineering
University of Salerno
Salerno, Italy
psommella@unisa.it

Avoci Ugwiri Moise
dept. of Industrial Engineering
University of Salerno
Salerno, Italy
mavociugwiri@unisa.it

Jan Lundgren
dept. Electronics Design
Mid Sweden University
Sundsvall, Sweden
jan.lundgren@miun.se

***Abstract*—Since early-stage skin cancer identification can improve melanoma prognosis and significantly reduce treatment costs, AI-based diagnosis systems might greatly benefit patients suffering from suspicious skin lesions. The study proposes a cosine cyclical learning rate with a skin cancer classification model to improve melanoma prediction. The contributions of models involve three critical CNNs, which are standard deep feature extraction modules for the skin cancer classification in this study (Vgg19, ResNet101 and InceptionV3). Each CNN model applies three different learning rates: fixed learning rate(LR), Cosine Annealing LR, and Cosine Annealing with WarmRestarts. HAM10000 is a large collection of publicly available dermoscopic images dataset used for our experiments. The performance of the proposed approach was appraised through comparative experiments. The outcome has indicated that the proposed method has high efficiency in diagnosing skin lesions with a cosine cyclical learning rate.**

***Index Terms*—cosine cyclical learning rate, deep learning, skin cancer, dermoscopic images, HAM10000**

I. Introduction

The vast majority of skin cancer deaths are from melanoma, one of the most common forms of cancer in the United States. Melanoma rates in the United States have risen rapidly over the past 40 years [1], shown in the following Fig. 1. Nearly 20 Americans die from melanoma every day [1]. In 2021, it is estimated that 7,180 deaths will be attributed to melanoma — 4,600 men and 2,580 women [2], [3]. The melanoma of the skin has 92% for 5-year relative survival from 2010 to 2016 [1], Which shows a high survival for early detection.

The diagnostic accuracy for two dermatologists, each with more than ten years of experience in dermatology, was 80%, with a sensitivity of 91% and a positive predictive value of 86%. Diagnostic accuracy rates for two senior registrars (each with 3-5 years of experience) and six registrars (each with 1-2 years of experience) were 62% and 56%, respectively [4]. The diagnostic accuracy of dermatologists is dependent on the degree of expertise of the examiners, and the costly equipment is required. Deep learning algorithms, driven by

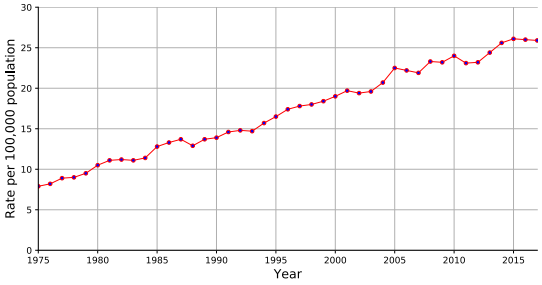


Fig. 1. The trends of Melanoma skin lesion in incidence rates from 1975 to 2017 in the US. The horizontal axis is the year and the vertical axis is per 100,000 population [1].

advances in computation and large datasets, have recently exceeded human performance in skin cancer diagnosis. In the last decade, convolution neural networks (CNNs) have been extensively utilized in skin lesion image analysis and have achieved remarkable results since they have robust feature representation capability [5]–[13]. The rapid development of artificial intelligence in skin diseases promises a more standardized level of diagnostic accuracy. Regardless of where they live or which doctor they see, all people can access reliable diagnostic assessments.

Esteva1 et al. [13] utilize a GoogleNet Inception v3 CNN architecture to train on their dataset of 129,450 skin lesions comprising to improve the accuracy of the classifiers for melanoma 2,032 different diseases and achieved the dermatologist-level classification of skin cancer. An ensemble of various CNNs also performed high classification accuracy shown by Harangi et al. [14] and Goyal et al. [15]. Zhao et al. [16] proposed a skin cancer classification framework based on a skin lesion augmentation style-based GAN with DenseNet201 to redesign the structure of style control and noise input and reconstruct the discriminator to adjust the generator to synthesize high-

978-1-6654-8360-5/22/$31.00 ©2022 IEEE

quality skin lesion images efficiently.

There is still much room for improvement in existing research on skin lesion diagnosis. No matter what algorithms are used, improving the performance of the model is still the goal that researchers are constantly pursuing.

II. Proposed Method

A. Simulated Annealing Algorithm

Deep learning optimization uses a greedy algorithm, a search-based optimization method that starts from point A when seeking the optimal solution. When seeking the optimal solution, it starts from point A, shown in Fig. 2. If the value of the function continues to decrease, then the test process continues. When point B is reached, the test process ends (no matter which direction it searches, The result will only get larger and larger), so the local optimal point B is found. When moving from B to the small peak between B and C, the probability of each shift to the right (that is, to accept a worse value) gradually decreases. If this slope is particularly long, then it is very likely that we will not climb this slope in the end. It is expected to be overturned if it is not too long, depending on the learning rate η value setting. If the calculated gradient in this direction is 0, the algorithm may stop iterating. When the learning rate is large, our gradient optimization process will be very violent and may reach the global minimum point. Still, it is also likely to be farther and farther away from the optimization goal. While the learning rate is minimal, we will drop smoothly, but it is easy to get stuck at the local minimum.

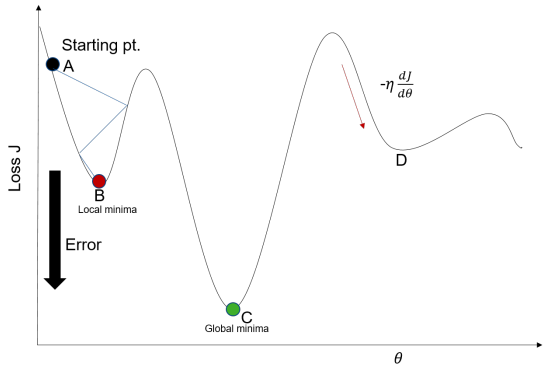


Fig. 2. Example of a gradient descent process.

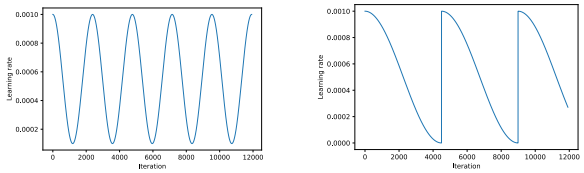
There are two problems with basic learning rate schedules : (1)we do not know what the optimal initial learning rate is; (2)monotonically decreasing our learning rate may lead to our network getting stuck in the plateaus of the loss landscape. Cyclical learning rates let the learning rate cyclically vary between reasonable boundary values. Training with cyclical learning rates instead of fixed one achieves improved classification accuracy. It does not require a tune and can reduce the number of iterations [17]. In other words, using Cyclical Learning Rates leads to faster convergence and with fewer experiments/hyperparameter updates.

This paper uses two cosine cyclical learning rates(see Fig. 3): the initial cosine cyclical learning rate, and another improved with warm restarts. Equation 1 [18] requires the total training epochs, maximum learning rate, and the number of cycles as arguments and the current epoch number. The function then returns the learning rate for the given epoch. The equation for the cosine annealing learning rate schedule Where α_t is the learning rate at epoch t, α_0 is the maximum learning rate, T is the total epochs.The training process will be split into M cycles. The mod is the modulo operation, and square brackets indicate a ceiling operation. The model converges to the first local minimum after a few epochs, then changes to a larger learning rate, jumps out of the current local, and then gradually reduces the learning rate, converges to the second local minimum. The model repeats this operation until the M^{th} local minimum. Because our objective optimization function may be multimodal (as shown in Fig. 2 below). In addition to the global optimal solution, there are multiple local optimal solutions. During training, the gradient descent algorithm may fall into a local minimum. At this time, it can be passed and suddenly increase the learning rate to leave the local minimum and find a path to the global minimum. This method is called the stochastic gradient descent method with a restart.

$$\alpha(t) = \frac{\alpha_0}{2}(cos(\frac{\pi mod(t-1, \lceil T/M \rceil)}{\lceil T/M \rceil}) + 1) \tag{1}$$

In PyTorch [19], the formula is also written like this:

$$\eta_t = \eta_{min} + \frac{1}{2}(\eta_{max} - \eta_{min})(1 + cos(\frac{T_{cur}}{T_{max}}\pi)) \tag{2}$$



(a) Cosine (b) Warm restarts

Fig. 3. The cosine cyclical learning rate and the cosine cyclical learning rate with warm restarts are based on iteration.

A similar cyclic approach is known for warm restarts, where an aggressive annealing schedule is combined with periodic restarts to the original starting learning rate, as shown in Fig. 3(b).

$$\eta_t = \eta_{min} + \frac{1}{2}(\eta_{max} - \eta_{min})(1 + cos(\frac{T_{cur}}{T_i}\pi)) \tag{3}$$

In a cosine annealing schedule, where η_t is the learning rate at timestep t. η_{max} and η_{min} define the range of desired learning rates. T_{cur} represents epochs since the last restart(this value is calculated at every iteration and thus can take on fractional values). T_{min} is the minimum learning rate and

T_{max} is the maximum number of iterations in Equation 2. T_i defines the number of epochs between two warm restarts in Equation 3.

B. Convolutional Models

The key CNN backbone architectures have been applied in this study. This study chooses the most popular pre-trained models here: Vgg19, ResNet101, Inception-V3.

- Vgg19. An improvement of VGGNet [20] over AlexNet [21] is to use several consecutive 3x3 convolution kernels to replace the larger convolution kernels in AlexNet (11x11, 5x5). For a given receptive field (the local size of the input image related to the output), using a stacked small convolution kernel is better than using a large convolution kernel, because multiple non-linear layers can increase the depth of the network to ensure a more complex learning mode, and the cost is relatively small (fewer parameters).
- ResNet101. ResNet was proposed by Kaiming et al. in 2015 [22]. ResNet designed a residual module to train a deeper network, allowing the gradient to pass smoothly in the network without exploding or disappearing. In this paper, we choose ResNet101 as one of our classifiers.
- Inception-V3. InceptionV3 [23] is an evolution of the original GoogleNet, where the 7×7 kernels are slipt into 1×7 and 7×1. This idea is also called Factorization into small convolutions thought. Moreover, it continues to reduce the top-5 error rate to 3.5% on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012 challenge. This network has a total of 42 layers, and its computation cost is about 2.5 times that GoogLeNet.

C. Datasets and Metrics

The HAM10000 dataset [24] consists of 10015 dermoscopic images, which include a representative collection of all important diagnostic categories in the realm of pigmented lesions: Actinic keratoses and intraepithelial carcinoma / Bowen's disease (akiec), basal cell carcinoma (bcc), benign keratosis-like lesions (solar lentigines / seborrheic keratoses and lichen-planus like keratoses, bkl), dermatofibroma (df), melanoma (mel), melanocytic nevi (nv) and vascular lesions (angiomas, angiokeratomas, pyogenic granulomas and hemorrhage, vasc).

There are four possibilities classification results: true positive, indicated by TP; false-positive, marked as FP; false-negative pointed out by FN, and true negative showed like TN. Based on these values, accuracy, precision, recall(sensitivity) and F1-score, these four measures are used to test the model robustness here:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$F1 - Score = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \tag{7}$$

Precision focuses on evaluating all the data predicted to be positive; how much does the real positive data account for. Recall focuses on evaluation: out of all the positive data, how many data have been successfully predicted as positive. The goal of the skin cancer diagnosis system is to increase the recall value as much as possible, even at the expense of some precision

Macro-average adds the evaluation indicators of different categories (precision/ recall/ F1-score) to average and gives all classes the same weight. This method can treat each class equally, but its value will be affected by the rare class.

Weighted-average gives different weights to different classes (the weight is determined according to the true distribution ratio of the classes). Each class is multiplied by the weight and then added. This method considers the imbalance of class, and its value is more susceptible to the influence of the majority class.

The meaning of AUC is the probability that when a positive sample and a negative sample are randomly selected, the positive sample will be ranked in front of the negative sample according to the score calculated by the current classifier. When AUC is 1, it is a perfect classifier.

III. Experimental Results and Discussion

A. Experimental Setup and Training

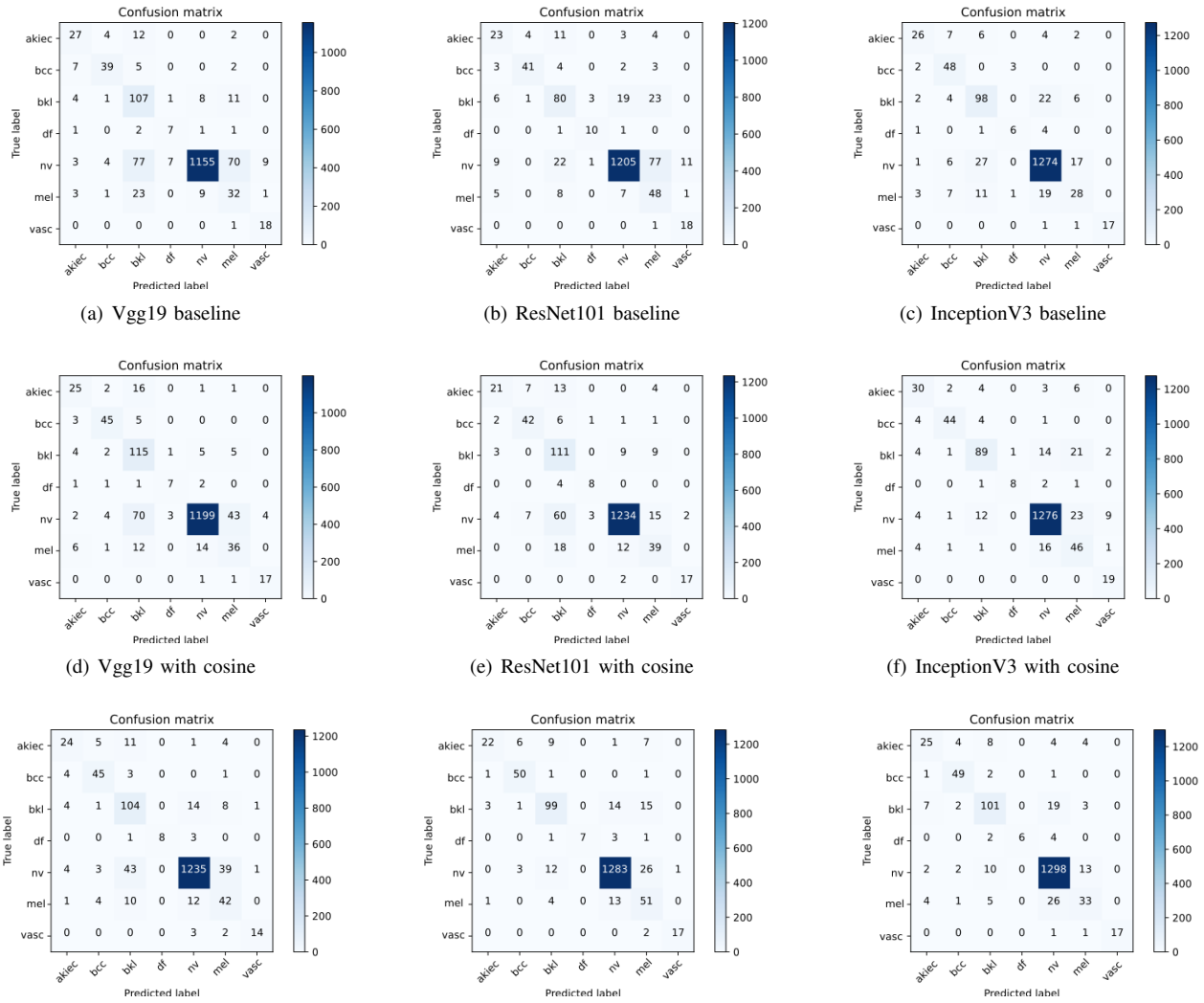
There are 4501 of the more than 10,000 images duplicates in the HAM10000 dataset. In the experiment, 30% of the 5514 images without repeated lesion id were classified as the validation set, and the 1655 pictures were stratified according to the lesion type as the label. In the training set, the number of lesion types is highly unbalanced. We use data augmentation: except for the nv class with the most significant number, all other categories are expanded by corresponding multiples in turn. The final training data is as follows Table I:

TABLE I
AUGMENT THE NUMBER OF UNBALANCED DATA IN HAM10000.

Class name	Raw data	Data enhancement
Melanoma	1044	6264
Basal cell carcinoma	461	5532
Vascular lesions	123	5412
Melanocytic nevi	5380	5380
Actinic keratoses	282	5358
Dermatofibroma	103	5356
Benign keratosis-like lesions	967	4835

This experiment adopts more data augmentations based on the HAM10000 datasets to generalize the model. We have employed several operations, including resizing images, random vertical and horizontal flipping, random rotation, and adjustments of color distributions.

Our experiments set the learning rate to 0.001 in the baseline model without changing. In the model, cosine annealing and



(a) Vgg19 baseline (b) ResNet101 baseline (c) InceptionV3 baseline

(d) Vgg19 with cosine (e) ResNet101 with cosine (f) InceptionV3 with cosine

(g) Vgg19 with warm (h) ResNet101 with warm (i) InceptionV3 with warm

Fig. 4. The nine confusion matrices of the nine classifiers' results.

cosine annealing with warm restarts, we set the initial learning rate as 0.001. All the models will take 1,192 iterations in each epoch. We change the learning rate of cosine periodically five times in cosine annealing models, as shown in Fig. 3(a). The models with warm restarts will restart the learning rate after four epochs shown in Fig. 3(b). The optimization under the guidance of our loss function cross-entropy is performed using the Adam optimizer [25]. Generally, our model can stably converge after ten epochs trained, where the batch size is 32. During the optimization, We also observe the learning rate of cyclical changes during the optimization in Fig. 3. We implemented the experiments using the PyTorch framework and a Linux system with GPU acceleration (Nvidia GeForce RTX 2080Ti graphics card).

B. Results of the Experiments and Discussion

When the dataset is balanced, accuracy is an effective metric to characterize the classification performance. However, HAM10000 is an imbalanced dataset, we evaluated the system with more performance metrics such as precision, recall and F1-score. We also applied a confusion matrix, which provides detailed information about correct and incorrect skin cancer classifications (in Fig. 4). While it is seen that all the models have the highest success in classifying nv. From Fig. 4, the performance of mel classification results with Vgg19 has been growing. It observes that the ResNet101 model improves the mel classification results in WarmRestarts but decreases in cosine. However, in the InceptionV3 model, the opposite is true.

For precision recall and F1-score, we calculate them separately for seven classes with nine different models given in Table II, III, IV. In multi-class (greater than two classes) skin cancer problems, accuracy measures the global sample prediction situation. We make a comparison of this study with nine different models. The experimental results are presented in Table V(the best performance is shown in bold). The highest accuracy we reached with ResNet101_warm

TABLE II
COMPARISON OF THE THREE DIFFERENT LEARNING RATES WITH VGG19

Lesion Class	Vgg19_base			Vgg19_cosine			Vgg19_warm		
	Precision	**Rrecall**	**F1-score**	**Precision**	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**
akiec	0.60	0.60	0.60	0.61	0.56	0.58	0.65	0.53	0.59
bcc	0.80	0.74	0.76	0.82	0.85	0.83	0.78	0.85	0.81
bkl	0.47	0.81	0.60	0.53	0.87	0.66	0.6	0.79	0.68
df	0.47	0.58	0.52	0.64	0.58	0.61	1.00	0.57	0.80
nv	0.98	0.87	0.92	0.98	0.90	0.94	0.97	0.93	0.95
mel	0.27	0.46	0.34	0.42	0.52	0.46	0.44	0.61	0.51
vasc	0.64	0.95	0.77	0.81	0.89	0.85	0.88	0.74	0.80

TABLE III
COMPARISON OF THE THREE DIFFERENT LEARNING RATES WITH RESNET101

Lesion Class	ResNet101_base			ResNet101_cosine			ResNet101_warm		
	Precision	**Rrecall**	**F1-score**	**Precision**	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**
akiec	0.50	0.51	0.51	0.70	0.47	0.56	0.81	0.49	0.61
bcc	0.89	0.77	0.83	0.75	0.79	0.77	0.83	0.94	0.88
bkl	0.63	0.61	0.62	0.52	0.84	0.65	0.79	0.75	0.77
df	0.71	0.83	0.77	0.67	0.67	0.67	1.00	0.58	0.74
nv	0.97	0.91	0.94	0.98	0.93	0.96	0.98	0.97	0.97
mel	0.31	0.70	0.43	0.57	0.57	0.57	0.5	0.74	0.59
vasc	0.60	0.95	0.73	0.89	0.89	0.89	0.94	0.89	0.92

TABLE IV
COMPARISON OF THE THREE DIFFERENT LEARNING RATES WITH INCEPTIONV3

Lesion Class	InceptionV3_base			InceptionV3_cosine			InceptionV3_warm		
	Precision	**Rrecall**	**F1-score**	**Precision**	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**
akiec	0.74	0.58	0.65	0.65	0.67	0.66	0.64	0.56	0.60
bcc	0.67	0.91	0.77	0.90	0.83	0.86	0.84	0.92	0.88
bkl	0.69	0.74	0.71	0.80	0.67	0.73	0.79	0.77	0.78
df	0.60	0.50	0.55	0.89	0.67	0.76	1.00	0.50	0.67
nv	0.96	0.96	0.96	0.97	0.96	0.97	0.96	0.98	0.97
mel	0.52	0.41	0.46	0.47	0.67	0.55	0.61	0.48	0.54
vasc	1.00	0.89	0.94	0.61	1.00	0.76	1.00	0.89	0.94

and InceptionV3_warm models is 92%. It can be seen that ResNet101_warm is the most efficient model, which also achieves the highest precision, F1-score and AUC. It is not appropriate to simply use macro average in an imbalanced dataset. At this time, weighted average can be taken. There is a big gap between macro average and weighted average in Table V.

Table V shows the effect of cosine cyclical learning rate on different CNN models. It can be found that after using cosine annealing and cosine annealing warm restarts, Vgg19 Accuracy increases 3% and 5%, respectively. ResNet101 and InceptionV3 have also improved significantly. Overall, cosine cyclical learning rate with deep learning can significantly improve the model's performance.

Our model generalization is insufficient. Since the ground truth of the ISIC 2018 Test dataset is not public, we did not test it on it. It is meaningless to compare the accuracy rate with other researchers only, so it is necessary to have an open unified standard test dataset. Our model is only tested on our own divided dataset. The comparison of this model on the unified test set proves the important impact of cosine cyclical learning rate on deep learning model training.

IV. CONCLUSIONS

Early detection is a crucial point for a successful cure of skin cancer. Our contribution to this paper is to use the cosine cyclical learning rate and verify its effectiveness with different CNN models. We have proposed several different CNN frameworks based on three different learning rates. In this study, CNN models work on the HAM10000 dataset with seven different types of skin lesion. Experiments showed that ResNet101 and InceptionV3 using cosine cyclical with warm restarts reached an accuracy of 92%, which are the best results. To compare with the baseline, all these three models (Vgg19, ResNet101 and InceptionV3) arrived at much better results. Therefore, the proposed models can efficiently improve the diagnosis performances of skin lesions. Skin lesion diagnosis based on deep learning offers the potential for earlier detection and treatment, which could improve survival. Further works will focus on the testing of more deep networks.

TABLE V
COMPARISON OF THE NINE MODELS WITH DIFFERENT AVERAGE METRICS

Models	Accuracy	Precision(M^1)	Precision(W^2)	Recall(M^1)	Recall(W^2)	F1-score(M^1)	F1-score(W^2)	AUC
Vgg19_base	0.84	0.60	0.89	0.72	0.84	0.64	0.86	0.96
Vgg19_cosine	0.87	0.69	0.90	0.74	0.87	0.70	0.88	0.97
Vgg19_warm	0.89	0.76	0.91	0.73	0.89	0.73	0.90	0.98
ResNet101_base	0.86	0.66	0.90	0.75	0.86	0.69	0.87	0.97
ResNet101_cosine	0.89	0.73	0.91	0.74	0.89	0.72	0.90	0.97
ResNet101_warm	**0.92**	**0.84**	**0.93**	0.77	**0.92**	**0.78**	**0.93**	**0.99**
InceptionV3_base	0.90	0.74	0.90	0.71	0.90	0.72	0.90	0.98
InceptionV3_cosine	0.91	0.76	0.92	**0.78**	0.91	0.76	0.92	0.98
InceptionV3_warm	**0.92**	**0.84**	0.92	0.73	**0.92**	0.77	0.92	0.98

[1] M: macro-average
[2] W: weighted-average

REFERENCES

[1] Cancer Statistics Center in the US., "Melanoma of the skin trends in incidence rates," https://cancerstatisticscenter.cancer.org/, 2021.

[2] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer statistics, 2021." *CA: a cancer journal for clinicians*, vol. 71, no. 1, pp. 7–33, 2021.

[3] "American cancer society: Cancer facts & figures 2021," *Atlanta: American Cancer Society*, 2021.

[4] C. Morton and R. Mackie, "Clinical accuracy of the diagnosis of cutaneous malignant melanoma." *The British journal of dermatology*, vol. 138, no. 2, pp. 283–287, 1998.

[5] Y. Wang, J. Cai, D. C. Louie, Z. J. Wang, and T. K. Lee, "Incorporating clinical knowledge with constrained classifier chain into a multimodal deep network for melanoma detection," *Computers in Biology and Medicine*, vol. 137, p. 104812, 2021.

[6] X. Wang, X. Jiang, H. Ding, Y. Zhao, and J. Liu, "Knowledge-aware deep framework for collaborative skin lesion segmentation and melanoma recognition," *Pattern Recognition*, p. 108075, 2021.

[7] R. Vani, J. Kavitha, and D. Subitha, "Novel approach for melanoma detection through iterative deep vector network," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–10, 2021.

[8] L. R. Soenksen, T. Kassis, S. T. Conover, B. Marti-Fuster, J. S. Birkenfeld, J. Tucker-Schwartz, A. Naseem, R. R. Stavert, C. C. Kim, M. M. Senna *et al.*, "Using deep learning for dermatologist-level detection of suspicious pigmented skin lesions from wide-field images," *Science Translational Medicine*, vol. 13, no. 581, 2021.

[9] R. C. Maron, S. Haggenmüller, C. von Kalle, J. S. Utikal, F. Meier, F. F. Gellrich, A. Hauschild, L. E. French, M. Schlaak, K. Ghoreschi *et al.*, "Robustness of convolutional neural networks in recognition of pigmented skin lesions," *European Journal of Cancer*, vol. 145, pp. 81–91, 2021.

[10] I. Bakkouri and K. Afdel, "Computer-aided diagnosis (cad) system based on multi-layer feature fusion network for skin lesion recognition in dermoscopy images," *Multimedia Tools and Applications*, vol. 79, no. 29, pp. 20 483–20 518, 2020.

[11] J. R. Hagerty, R. J. Stanley, H. A. Almubarak, N. Lama, R. Kasmi, P. Guo, R. J. Drugge, H. S. Rabinovitz, M. Oliviero, and W. V. Stoecker, "Deep learning and handcrafted method fusion: higher diagnostic accuracy for melanoma dermoscopy images," *IEEE journal of biomedical and health informatics*, vol. 23, no. 4, pp. 1385–1391, 2019.

[12] S. S. Han, M. S. Kim, W. Lim, G. H. Park, I. Park, and S. E. Chang, "Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm," *Journal of Investigative Dermatology*, vol. 138, no. 7, pp. 1529–1538, 2018.

[13] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[14] B. Harangi, "Skin lesion classification with ensembles of deep convolutional neural networks," *Journal of biomedical informatics*, vol. 86, pp. 25–32, 2018.

[15] M. Goyal, A. Oakley, P. Bansal, D. Dancey, and M. H. Yap, "Skin lesion segmentation in dermoscopic images with ensemble deep learning methods," *IEEE Access*, vol. 8, pp. 4171–4181, 2019.

[16] C. Zhao, R. Shuai, L. Ma, W. Liu, D. Hu, and M. Wu, "Dermoscopy image classification based on stylegan and densenet201," *IEEE Access*, vol. 9, pp. 8659–8679, 2021.

[17] L. N. Smith, "Cyclical learning rates for training neural networks," in *2017 IEEE winter conference on applications of computer vision (WACV)*. IEEE, 2017, pp. 464–472.

[18] G. Huang, Y. Li, G. Pleiss, Z. Liu, J. E. Hopcroft, and K. Q. Weinberger, "Snapshot ensembles: Train 1, get m for free," *arXiv preprint arXiv:1704.00109*, 2017.

[19] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," 2017.

[20] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[23] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[24] O. Sevli, "A deep convolutional neural network-based pigmented skin lesion classification application and experts evaluation," *Neural Computing and Applications*, pp. 1–12, 2021.

[25] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.