BOOSTING MEDICAL IMAGE CLASSIFICATION WITH SEGMENTATION FOUNDATION MODEL

Pengfei Gu[†] *Zihan Zhao*[◇] *Hongxiao Wang*[†] *Yaopeng Peng*[†] *Yizhe Zhang*[⋆1] *Nishchal Sapkota*[†]
Chaoli Wang[†] *Danny Z. Chen*[†]

[⋆]Nanjing University of Science and Technology, Nanjing, Jiangsu 210094, China
[†]University of Notre Dame, Notre Dame, IN 46556, USA
[◇]Tianjin University, Tianjin, Tianjin 300072, China

ABSTRACT

The Segment Anything Model (SAM) exhibits impressive capabilities in zero-shot segmentation for natural images. Recently, SAM has gained a great deal of attention for its applications in medical image segmentation. However, to our best knowledge, no studies have shown how to harness the power of SAM for medical image classification. To fill this gap and make SAM a true "foundation model" for medical image analysis, it is highly desirable to customize SAM specifically for medical image classification. In this paper, we introduce SAMAug-C, an innovative augmentation method based on SAM for augmenting classification datasets by generating variants of the original images. The augmented datasets can be used to train a deep learning classification model, thereby boosting the classification performance. Furthermore, we propose a novel framework that simultaneously processes raw and SAMAug-C augmented image input, capitalizing on the complementary information that is offered by both. Experiments on three public datasets validate the effectiveness of our new approach.

1. INTRODUCTION

Trained on over 1 billion tasks using 11 million images, the Segment Anything Model (SAM) [1], a Segmentation Foundation Model, has showcased impressive zero-shot image segmentation capabilities for natural images across various prompts, such as points, boxes, and masks. Recently, a number of studies have explored leveraging SAM for medical image segmentation, either by directly applying SAM [2, 3, 4, 5] or by fine-tuning SAM for medical images [6, 7, 8]. Despite these efforts, some studies have exhibited unsatisfactory performance of using SAM in medical image segmentation [6, 2] due to the following challenges: (1) the large differences in appearance between medical and natural images, and (2) the often blurred boundaries of target objects in medical images. In contrast to the SAM-based methods mentioned above for medical image segmentation, SAMAug [5] employed SAM to augment raw image input for commonly-used deep learning (DL) medical image segmentation models (e.g., U-Net [10]), thereby enhancing their segmentation performances. Specifically, SAMAug augments raw images with segmentation maps and boundary prior maps generated by SAM (i.e., adding the segmentation and boundary prior maps to the second and third channels of the raw images, respectively).

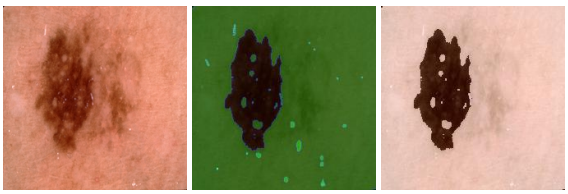

Fig. 1. Visual example from the ISIC 2017 dataset [9]. The entire image augmented by SAMAug [5] is covered in green, which may increase the difficulty for a classifier to distinguish the skin lesion from the background.

Medical image classification is a pivotal task in diagnostic medicine, assisting clinicians in their decision-making processes [11]. To our best knowledge, SAM has yet to be employed for medical image classification. While one might consider using SAMAug, inspired by its capabilities, to augment raw image input for classification, several challenges arise. First, SAMAug was primarily designed to augment raw images for medical image segmentation, emphasizing the use of both SAM-generated segmentation maps and boundary prior maps. Our experiments have shown that applying SAMAug directly to raw images intended for classification could lead to a performance drop. As evidence, in Table 1, we observe performance drops on the ISIC 2017 skin lesion classification dataset [9] when training two DL-based classification models (ResNet152 [12] and SENet154 [13]) with raw images augmented by SAMAug. Second, the SAM-generated segmentation maps and boundary prior maps could inadvertently obscure crucial regions in the raw images. Visual example of such issues with augmented image produced

[1] Corresponding author

Table 1. Results on the ISIC 2017 dataset.

Method	Acc (↑)	AUC (↑)	Sen (↑)	Spe (↑)
ResNet152 [12]	**84.53**	**81.28**	**49.23**	**93.08**
ResNet152 + SAMAug [5]	82.13	76.01	37.78	92.88
SENet154 [13]	**84.45**	**79.41**	**42.74**	**94.53**
SENet154 + SAMAug [5]	82.53	77.00	37.09	93.54

by SAMAug, using sample from the ISIC 2017 skin lesion classification dataset, is illustrated in Fig. 1.

With these challenges in mind, it becomes imperative to customize SAM specifically for medical image classification. This paper addresses two pivotal questions for medical image classification: (I) How can we design an effective SAM-based augmentation method that emphasizes the crucial regions while suppressing irrelevant ones in input images for medical image classification? (II) How can we effectively utilize raw and SAM-augmented images to enhance classification performance?

For the first question, we introduce SAMAug-C, an innovative augmentation method built upon SAM, aiming to augment the input datasets by creating variations of the original images. Initially, we leverage SAM's zero-shot image segmentation capability to procure segmentation masks for the raw images. We then generate the corresponding segmentation prior maps by assigning a value of '1' to the masked regions. Subsequently, these segmentation prior maps are added to each channel of the raw images to augment them.

For the second question, we propose a novel framework that processes both the raw and SAMAug-C augmented images simultaneously, harnessing the complementary information that each of them provides. The framework consists of two branches, both equipped with identical backbone models. These models are concurrently trained with the raw and SAMAug-C augmented images. Subsequently, an ensemble module is employed to amalgamate the predictions from the two branches, yielding the final predicted label.

In summary, our main contributions are as follows: (1) We adapt SAM specifically for medical image classification, pioneering its use in this domain. (2) We present SAMAug-C, which is designed to augment raw images for medical image classification, and put forward a new framework that effectively trains on both raw and SAMAug-C augmented images simultaneously. (3) We conduct comprehensive experiments on three public medical image classification datasets to demonstrate the effectiveness of our new approach.

2. METHODOLOGY

2.1. Background: The SAM Architecture

SAM consists of three main components: an image encoder, a prompt encoder, and a mask decoder. The image encoder accepts an input image of any size and produces an embedding feature. The prompt encoder can handle both sparse prompts (e.g., boxes) and dense prompts (e.g., masks). The mask decoder is a Transformer decoder block modified to incorporate a dynamic mask prediction head. SAM employs a two-way attention module, one for prompt-to-image embedding and the other for image-to-prompt embedding in each block, facilitating learning of the interactions between the prompt and image embeddings. After processing through two blocks, SAM upsamples the image embedding. Then, a MLP maps the output token to a dynamic linear classifier, which then predicts the target mask for the provided image. In this paper, we leverage SAM to predict the mask of the input image.

2.2. SAMAug-C: Augmenting Input Raw Images for Medical Image Classification

For a given raw image I, SAMAug [5] produces two corresponding prior maps for I. The first one is a segmentation prior map derived from the mask's stability score generated by SAM. The second one is a boundary prior map, representing the exterior boundary of the segmentation mask. The raw image I is augmented by adding to I the segmentation prior map to its second channel and the boundary prior map to its third channel. While this augmentation strategy was shown to be effective for several medical image segmentation tasks [5], it falls short in performance for medical image classification (e.g., as shown in Table 1).

To address this limitation and develop a more effective augmentation method for highlighting the important regions and suppressing irrelevant ones in the input images for medical image classification, we introduce SAMAug-C, augmenting raw image input for medical image classification.

As illustrated in Algorithm 1, for a given input raw image, SAM's mask generator first predicts segmentation masks and stores them in a list. For every segmentation mask in this list, we generate a corresponding segmentation prior map, assigning a value of 1 to the masked region. Then, we combine all the segmentation prior maps to produce a final segmentation prior map for the raw image. We then set the values of all the masked regions in this final segmentation prior map to 1, and augment the raw image by adding the segmentation prior map to each channel of the raw image. It is important to note that if SAM does not generate any segmentation masks, then the raw image remains unaugmented. The added segmentation prior map effectively emphasizes crucial regions and diminishes irrelevant ones in the input image. Refer to Fig. 1 for visual example of image augmented using SAMAug-C.

2.3. Model Training with Raw and SAMAug-C Augmented Images

Using SAMAug-C to augment input raw images, we derive a new set of images, called SAMAug-C augmented images. This raises a pertinent question: How can we effectively leverage the SAMAug-C augmented images to enhance medical image classification? A straightforward approach may involve: (I) During the training phase, employ SAMAug-C augmented images to train a DL classification model; (II) in the test phase, labels are predicted by feeding the model with SAMAug-C augmented test images. We refer to this approach as "DL classification model + SAMAug-C" (e.g., ResNet152 + SAMAug-C). While possibly effective, this

Algorithm 1 SAMAug-C($tI, mask\_generator$)

Input: The raw image input tI and mask generator $mask\_generator$ from SAM
Output: Augmented image $newTI$
$masks \leftarrow mask\_generator$.GENERATE($tI$)
$tI \leftarrow$ IMG_AS_FLOAT(tI)
$SegPrior \leftarrow$ ZEROSMATRIX(size of $tI[0], tI[1]$)
$newTI \leftarrow$ ZEROSMATRIX(size of $tI[0], tI[1], tI[2]$)
for $maskindex$ **from** 0 **to** length of $masks - 1$ **do**
    $thismask \leftarrow masks[maskindex]['segmentation']$
    $thismask\_ \leftarrow$ ZEROSMATRIX(size of $thismask$)
    $thismask\_$[where $thismask$ = True] $\leftarrow 1$
    $SegPrior$[where $thismask\_ = 1$] $\leftarrow$ $SegPrior$[where $thismask\_ = 1$] $+ 1.0$
if $SegPrior$.MIN $=$ $SegPrior$.MAX **then**
    for i **from** 0 **to** 2 **do**
        $newTI[:, :, i] \leftarrow tI[:, :, i]$
else
    $SegPrior \leftarrow$ WHERE($SegPrior \neq 0, 1, 0$)
    for i **from** 0 **to** 2 **do**
        $newTI[:, :, i] \leftarrow tI[:, :, i] + SegPrior$
return $newTI$

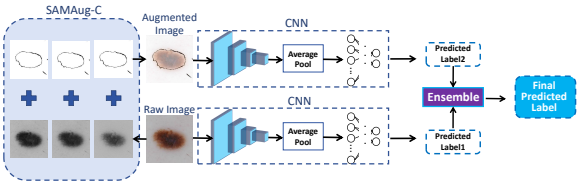

Fig. 2. The overview of our framework.

straightforward design might struggle in scenarios where SAM fails to produce accurate segmentation masks.

To address this limitation, we propose a novel framework that concurrently processes both the raw images and their SAMAug-C augmented images. As shown in Fig. 2, our new framework consists of two branches. The "Raw Image" branch ingests raw image input to train a DL classification model (e.g., ResNet) and produces predicted labels. In contrast, the "Augmented Image" branch uses the augmented images generated by SAMAug-C to train another DL classification model, which shares the same architecture as that of the "Raw Image" branch, and subsequently outputs the associated predicted labels. An ensemble module is employed to consolidate the outputs of the two branches, generating the final predicted label. We refer to this design as "DL classification model + SAMAug-C + Ensemble" (e.g., ResNet152 + SAMAug-C + Ensemble).

2.4. Model Ensemble

For a given test image x and its corresponding augmented test image generated by SAMAug-C, we employ two models to process them and ensemble the results. From these two models, we obtain two N-dimensional vectors, p_1 and p_2, where N denotes the number of categories that the task is classified into. We explore several possible ensemble schemes: **(1) Voting**: The majority voting strategy is employed to derive the final prediction by consolidating the outcomes from both models. **(2) Entropy**: The entropy offers a quantifiable measure of the uncertainty or randomness present in the predicted probabilities across various classes. We compute the entropy using the equation $-p \times \log_2 p$, where p is the set of predicted probabilities. By multiplying each probability with its logarithm (base 2) and negating the result, we then accumulate these values across the classes to determine the entropy for each sample. Ultimately, the model output with the lower entropy is chosen from the two. **(3) Direct average**: The outcomes from both models are averaged to arrive at the final prediction. **(4) Weighted average**: During the averaging process, a weight, indicating the significance of each model's output, is employed: Final prediction $= \frac{\sum_{i=1}^{2} \omega_i p_i}{2}$, where ω_i represents the weight of the output of the i-th model, with $\omega_i \geq 0$ and $\omega_1 + \omega_2 = 1$. For clarity, we refer to the "Raw Image" branch as the 1-st model and the "Augmented Image" branch as the 2-nd model.

Through empirical study, we choose the weighted average ensemble scheme for our framework in all the experiments.

3. EXPERIMENTS AND RESULTS

Datasets. (1) The ISIC 2017 skin lesion classification dataset (ISIC 2017):The dataset [9] contains 2000 training, 150 validation, and 600 test images. Our experiments are focused on task-3A: melanoma detection. (2) The vitiligo (public) dataset:The dataset [14] contains 672 training, 268 validation, and 401 test images. Our experiments are performed for vitiligo detection. (3) The extended colorectal cancer (ExtCRC) grading dataset: The dataset [15] contains 300 H&E-stained colorectal cancer subtyping pathology images. The task is three categories (Grades 1, 2, and 3) classification. We randomly split the data, allocating 80% for training and 20% for testing. We resize all the images of each dataset to 224×224. For the ISIC 2017 and vitiligo public datasets, our experiments conduct 5 runs using different seeds, and for the ExtCRC dataset, we perform the random data splitting for 5 times, presenting the average outcomes of our experiments.

Implementation Details. Our experiments are conducted using the PyTorch. The model is trained on an NVIDIA Tesla V100 Graphics Card (32GB GPU memory) using the AdamW optimizer with a weight decay = 0.005. The learning rate is 0.0001, and the number of training epochs is 400 for the experiments. The batch size for each case is set as the maximum size allowed by the GPU. Standard data augmentation (e.g., random flip, crop, etc.) is applied to avoid overfitting.

Experimental Results. (1) ISIC 2017 Results. To evaluate our method, we use two prominent models: ResNet152 and SENet154, which were pre-trained on ImageNet and have proven their efficacy on medical image datasets, with teams using them to obtain top scores in the ISIC skin lesion classification challenge. From Table 2, we observe that: (I) The mod-

Table 2. Results on the ISIC 2017 dataset. The best results are marked in **bold**, and the second-best results are underlined. Same for the other tables.

Method	Acc (↑)	AUC (↑)	Sen (↑)	Spe (↑)
Galdran et al. [16]	48.00	76.50	**90.60**	37.70
Vasconcelos et al. [17]	83.00	79.10	17.10	99.00
Díaz [18]	82.30	**85.60**	10.30	**99.80**
Zhang et al. [19]	83.00	83.00	—	—
Suraj et al. [20]	86.00	83.10	59.00	92.60
ResNet152 [12]	84.53	81.28	49.23	93.08
SENet154 [13]	84.45	79.41	42.74	94.53
ResNet152 + SAMAug-C	85.07	81.83	48.89	94.20
ResNet152 + SAMAug-C + Ensemble	85.67	83.93	48.72	94.62
SENet154 + SAMAug-C	85.83	80.82	44.79	95.78
SENet154 + SAMAug-C + Ensemble	**86.67**	83.27	47.01	96.27

Table 3. Results on the vitiligo (public) dataset.

Method	Acc (↑)	AUC (↑)	Sen(↑)	Spe (↑)
VGG13 [21]	—	0.995	0.972	0.963
ResNet18 [12]	—	0.958	0.952	0.957
DenseNet121 [22]	—	0.982	0.962	0.961
Dermatologists [14]	—	—	0.964	0.803
Suraj et al. [20]	0.988	**0.998**	0.996	**0.975**
ResNet18 + SAMAug [5]	0.967	0.992	0.990	0.934
DenseNet121 + SAMAug [5]	0.975	0.993	0.983	0.957
ResNet18 + SAMAug-C	0.968	0.993	0.991	0.955
ResNet18 + SAMAug-C + Ensemble	0.982	**0.998**	0.996	0.966
DenseNet121 + SAMAug-C	0.978	0.993	0.993	0.962
DenseNet121 + SAMAug-C + Ensemble	**0.990**	**0.998**	**0.997**	**0.975**

els trained with SAMAug-C augmented images demonstrate superior performance over their counterparts trained solely on raw images. Specifically, ResNet152's accuracy is improved by 0.54%, while SENet154's is improved by 1.38%. This validates the effectiveness of our SAMAug-C augmentation method in boosting medical image classification. (II) When the models are concurrently trained with both raw and SAMAug-C augmented images, their performances are further bolstered. ResNet152's accuracy is lifted by an additional 0.6%, while SENet154's grows by 0.84%. This finding shows our proposed framework's capability to leverage raw and SAMAug-C augmented images to elevate classification outcomes. (III) Our method outperforms the SOTA methods (Suraj et al. [20] and Zhang et al. [19]) in accuracy. **(2) Vitiligo (Public) Results.** From Table 3, we observe that: (I) On employing SAMAug-C augmented images for training, both our baseline models, ResNet18 and DenseNet121, exhibit superior results. Specifically, there is an improvement in AUC by 3.50% with ResNet18 and 1.10% with DenseNet121. This further attests to the capability of our SAMAug-C augmentation method. (II) When the models are simultaneously trained on raw and SAMAug-C augmented images, the AUC results for ResNet18 and DenseNet121 attain an uplift by an additional 0.50%. This indicates that our dual training approach can enhance performance over using just augmented or raw images alone. (III) Our method slightly outperforms the SOTA method, Suraj et al. [20], in both accuracy and sensitivity. These results demonstrate our method's effectiveness. **(3) ExtCRC Results.** In the experiments, we use non-pre-trained versions of three representative (ResNet50, ResNeXt50, and SE-ResNet50) models for gauging the robustness and stability of our method. From Table 4, we observe that: (I) Augmenting raw images using our SAMAug-C method leads to a noticeable performance improvement for

Table 4. Results on the ExtCRC dataset.

Method	Acc (↑)	AUC (↑)	Sen (↑)	Spe (↑)
ResNet50 [12]	77.00	89.07	74.06	87.93
ResNeXt50 [23]	80.00	90.93	77.28	89.40
SE-ResNet50 [13]	81.33	90.85	78.76	90.36
ResNet50 + SAMAug [5]	76.67	88.73	71.63	87.52
ResNeXt50 + SAMAug [5]	76.00	88.64	71.67	87.16
SE-ResNet50 + SAMAug [5]	81.25	90.71	77.44	89.86
ResNet50 + SAMAug-C	79.67	90.85	76.92	89.41
ResNet50 + SAMAug-C + Ensemble	80.67	91.68	78.06	89.70
ResNeXt50 + SAMAug-C	81.67	90.80	79.19	90.42
ResNeXt50 + SAMAug-C + Ensemble	82.46	91.67	80.62	90.87
SE-ResNet50 + SAMAug-C	82.67	92.02	81.06	91.19
SE-ResNet50 + SAMAug-C + Ensemble	**83.52**	**92.69**	**82.18**	**91.82**

Table 5. Results of various ensemble methods on the ISIC 2017 dataset.

	Acc (↑)	AUC (↑)	Sen (↑)	Spe (↑)
ResNet152 + SAMAug-C				
Voting	84.83	82.29	49.57	93.37
Entropy	82.67	81.89	38.46	93.37
Direct Average	85.17	83.33	50.43	93.58
Weighted Average w/ weights [0.6, 0.4]	85.33	84.18	50.43	93.79
Weighted Average w/ weights [0.4, 0.6]	84.67	81.15	55.56	91.72
Weighted Average w/ weights [0.7, 0.3]	84.17	**84.30**	**59.83**	90.06
Weighted Average w/ weights [0.3, 0.7]	**85.67**	83.93	48.72	**94.62**
SENet154 + SAMAug-C				
Voting	86.00	80.78	**48.72**	95.03
Entropy	84.84	82.18	45.30	94.41
Direct Average	86.50	83.30	47.01	96.07
Weighted Average w/ weights [0.6, 0.4]	86.33	81.05	47.86	95.65
Weighted Average w/ weights [0.4, 0.6]	86.50	83.35	47.86	95.86
Weighted Average w/ weights [0.7, 0.3]	86.17	81.12	**48.72**	95.24
Weighted Average w/ weights [0.3, 0.7]	**86.67**	**83.37**	47.01	**96.27**

all three non-pre-trained models. Interestingly, when these models are trained using augmented raw images generated by the SAMAug [5], a decline in their performance is evident. This observation highlights the superiority and appropriateness of our SAMAug-C augmentation method for medical image classification. (II) We observe further improvements in model accuracy by simultaneously training on both raw images and their SAMAug-C augmented images. Concretely, ResNet50, ResNeXt50, and SE-ResNet50 are improved by 1.0%, 0.79%, and 0.85% in accuracy, respectively. These results are a testimony to the potential of our method.

Model Ensemble Exploration. We conduct experiments to explore different ensemble methods (i.e., voting, entropy, direct average, and weighted average with various weights) using the ISIC 2017 dataset. As shown in Table 5, for both ResNet152 and SENet154, the weighted average method yields the best accuracy results. The direct average method gives the second-best results, followed by voting. The entropy method attains the least accurate results. The combination of weights [0.3, 0.7] produces the highest accuracy among the different weight choices for the weighted average method. This suggests that predictions by the model trained with SAMAug-C augmented images are more reliable.

4. CONCLUSIONS

In this paper, we presented a new augmentation method (SAMAug-C) that leverages the SAM for augmenting raw image input to improve medical image classification. To further enhance classification performance, we designed a novel framework that effectively uses both raw and SAMAug-C augmented images. Experiments on three public datasets demonstrated the efficacy of our new approach.

5. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using human subject data made available in open access by three publicly available datasets [9, 14, 15]. Ethical approval was not required as confirmed by the licenses attached with the open access datasets.

6. ACKNOWLEDGEMENTS

This research was supported in part by NSF grants IIS-1955395, IIS-2101696, and OAC-2104158.

7. REFERENCES

[1] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. B. Girshick, "Segment anything," *arXiv preprint arXiv:2304.02643*, 2023.

[2] R. Deng, C. Cui, Q. Liu, T. Yao, L. W. Remedios, S. Bao, B. A. Landman, L. E. Wheless, L. A. Coburn, K. T. Wilson, Y. Wang, S. Zhao, A. B. Fogo, H. Yang, Y. Tang, and Y. Huo, "Segment anything model (SAM) for digital pathology: Assess zero-shot segmentation on whole slide imaging," *arXiv preprint arXiv:2304.04155*, 2023.

[3] C. Mattjie, L. V. de Moura, R. C. Ravazio, L. S. Kupssinskü, O. Parraga, M. M. Delucis, and R. C. Barros, "Exploring the zero-shot capabilities of the segment anything model (SAM) in 2D medical imaging: A comprehensive evaluation and practical guideline," *arXiv preprint arXiv:2305.00109*, 2023.

[4] Y. Huang, X. Yang, L. Liu, H. Zhou, A. Chang, X. Zhou, R. Chen, J. Yu, J. Chen, C. Chen, H. C. Chi, X. Hu, D. Fan, F. Dong, and D. Ni, "Segment anything model for medical images?," *arXiv preprint arXiv:2304.14660*, 2023.

[5] Y. Zhang, T. Zhou, P. Liang, and D. Z. Chen, "Input augmentation with SAM: Boosting medical image segmentation with segmentation foundation model," *arXiv preprint arXiv:2304.11332*, 2023.

[6] J. Ma and B. Wang, "Segment anything in medical images," *arXiv preprint arXiv:2304.12306*, 2023.

[7] Y. Li, M. Hu, and X. Yang, "Polyp-SAM: Transfer SAM for polyp segmentation," *arXiv preprint arXiv:2305.00293*, 2023.

[8] K. Zhang and D. Liu, "Customized segment anything model for medical image segmentation," *arXiv preprint arXiv:2304.13785*, 2023.

[9] N. C. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," in *ISBI*, 2018, pp. 168–172.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *MICCAI*. 2015, vol. 9351, pp. 234–241, Springer.

[11] R. J. Chen, M. Y. Lu, J. Wang, D. F. Williamson, S. J. Rodig, N. I. Lindeman, and F. Mahmood, "Pathomic fusion: An integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis," *IEEE Transactions on Medical Imaging*, vol. 41, no. 4, pp. 757–770, 2020.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, 2016, pp. 770–778.

[13] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *CVPR*, 2018, pp. 7132–7141.

[14] L. Zhang, S. Mishra, T. Zhang, Y. Zhang, D. Zhang, Y. Lv, M. Lv, N. Guan, X. S. Hu, D. Z. Chen, and X. Han, "Design and assessment of convolutional neural network based methods for vitiligo diagnosis," *Frontiers in Medicine*, vol. 8, pp. 754202, 2021.

[15] M. Shaban, R. Awan, M. M. Fraz, A. Azam, Y.-W. Tsang, D. Snead, and N. M. Rajpoot, "Context-aware convolutional neural network for grading of colorectal cancer histology images," *IEEE Transactions on Medical Imaging*, vol. 39, no. 7, pp. 2395–2405, 2020.

[16] A. Galdran, A. Alvarez-Gila, M. I. Meyer, C. L. Saratxaga, T. Araújo, E. Garrote, G. Aresta, P. Costa, A. M. Mendonça, and A. Campilho, "Data-driven color augmentation techniques for deep skin image analysis," *arXiv preprint arXiv:1703.03702*, 2017.

[17] C. N. Vasconcelos and B. N. Vasconcelos, "Increasing deep learning melanoma classification by classical and expert knowledge based image transforms," *arXiv preprint arXiv:1702.07025*, 2017.

[18] I. G. Díaz, "Incorporating the knowledge of dermatologists to convolutional neural networks for the diagnosis of skin lesions," *arXiv preprint arXiv:1703.01976*, 2017.

[19] J. Zhang, Y. Xie, Q. Wu, and Y. Xia, "Medical image classification using synergic deep learning," *Medical Image Analysis*, vol. 54, pp. 10–19, 2019.

[20] S. Mishra, Y. Zhang, L. Zhang, T. Zhang, X. S. Hu, and D. Z. Chen, "Data-driven deep supervision for skin lesion classification," in *MICCAI*, 2022, pp. 721–731.

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

[22] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *CVPR*, 2017, pp. 4700–4708.

[23] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *CVPR*, 2017, pp. 1492–1500.