

Anti-Occlusion Diagnosis of Skin Cancer Based on Heterogeneous Data

Zhecheng Wu^{1,2}, Lu Leng^{1,2*}

Abstract

Skin cancer is a major disease seriously threatening human health, so it is significant to detect and diagnose it as early as possible. An automatic diagnosis model is constructed to predict the probability that a skin cancer case is benign or malignant. ISIC 2024 dataset is used, which has heterogeneous data, including structured data and unstructured data. Structured data include metadata and attribute values, which are mined and cleaned to generate more helpful data. Unstructured data are the images, many of which are occluded. SelfClean technique is utilized to select the occluded images that are then repaired using an improved SS-MAT algorithm. SS-MAT algorithm is enhanced by attention mechanism and a neural network, so it is more suitable for the spatial continuity of skin cancer. Deep learning models are used to extract features from the repaired images. The features extracted from structured data and unstructured data are inputted to three classifiers, namely LightGBM, CatBoost and XGBoost, which vote jointly to make the final decision. The custom metric score is improved from 0.1740 to 0.1935, which significantly improves the prediction accuracy of skin cancer.

Key Words: Medical Images, Image Repair, Anti-Occlusion Ability, Medical Diagnosis, Skin Cancer.

I. INTRODUCTION

Skin cancer, as a major disease seriously threatening human health [1], its early detection and diagnosis play a decisive role [2]. In the medical field, early diagnosis of skin cancer can significantly improve the cure rate and survival quality [3]. There is a severe shortage of specialized dermatologists in many areas, especially those with relatively scarce medical resources [4]. A large number of skin cancer patients can not be diagnosed and treated early, and their conditions are accordingly delayed [5].

ISIC 2024-Skin Cancer Detection with 3D-TBP dataset [6] is a popular skin cancer dataset, which has heterogeneous data, namely structured data and unstructured data. Structured data include metadata and attribute values; while unstructured data are composed of the images. The metadata cover some important information, as shown in Table 1, including the data source, patients' diagnostic information, symptom manifestations, etc. Data source helps researchers understand and analyze the differences between the data from different sources; the patients' diagnostic in-

formation includes the patients' past medical history; symptom manifestations provide a rich reference for comprehensively evaluating the patients' conditions.

There are three challenges for the automatic diagnosis of skin cancer based on heterogeneous data. The first one is that the raw structured data cannot sufficiently employed. The second one is that many images are occluded, as shown in Fig. 1. The third one is that it is difficult fuse the heterogeneous data to achieve their complementary advantages. To address the two problems, the contributions of this paper can be summarized as follows.

(1) The structured data are mined and cleaned to generate more helpful data. 71 optimized attributes are constructed or selected for diagnosis by removing the useless attributes and the redundancy.

(2) Many images, as unstructured data, are occluded by the patients' own hair, clothing, and jewelry, or due to shooting angle and distance. SelfClean [7] technique can accurately identify occluded images. Squeeze-and-Excitation Networks SIREN Mask-Aware Transformer (SS-MAT)

Manuscript received April 25, 2025; Revised May 02, 2025; Accepted May 04, 2025. (ID No. JMIS-25M-04-009)

Corresponding Author (*): Lu Leng, +86-791-86453251, leng@nchu.edu.cn

¹Jiangxi Provincial Key Laboratory of Image Processing and Pattern Recognition, Nanchang Hangkong University, Nanchang 330063, China, wzc1002235973@163.com, leng@nchu.edu.cn

²School of Software, Nanchang Hangkong University, 330063, China. wzc1002235973@163.com, leng@nchu.edu.cn

Table 1. Structured data description of ISIC 2024 dataset.

Attribute	Description
ISIC_id	Unique lesion identifier
patient_id	Unique patient identifier
age_approx	Age of patient
sex	Sex of patient
anatom_site_general	Location of the lesion on a patient's body
clin_size_long_diam_mm	Maximum diameter of the lesion
image_type	Structured field of the ISIC Archive for image type
tbp_tile_type	Lighting modality of the 3D TBP source image
tbp_lv_A	A inside lesion
tbp_lv_Aex	A outside lesion
tbp_lv_B	B inside lesion
:	:
tbp_lv_symm_2axis_angle	Lesion border asymmetry angle
tbp_lv_x	X-coordinate of the lesion on 3D TBP
tbp_lv_y	Y-coordinate of the lesion on 3D TBP
tbp_lv_z	Z-coordinate of the lesion on 3D TBP



Fig. 1. Occlude images.

is used to repair the occluded images, which is enhanced by attention mechanism and a neural network, so it is more suitable for the spatial continuity of skin cancer. Then deep learning models are used to extract the features from the repaired images.

(3) Finally the features extracted from structured data and unstructured data are inputted to three classifiers, namely LightGBM, CatBoost and XGBoost, which vote jointly to make the final decision.

The score of the model is improved from 0.1740 to 0.1935, which significantly improves the prediction accuracy of skin cancer.

II. RELATED WORKS

In the biomedical field, disease prediction faces many challenges [8], including the high complexity of the human physiological system, the diversity of disease triggers, and the significant differences between individuals. It requires not only processing huge amounts of biomedical data [9],

but also accurately understanding the intrinsic laws of disease development.

In order to overcome these challenges, machine learning algorithms are commonly used and combined to build reliable models for disease prediction [10]. In diabetes prediction research, Choudhury et al. [11] attempted to apply machine learning to diabetes prediction in 2019. Yang et al. [12] proposed the StoolNet model for color classification of fecal medical images in 2019. In 2022, Leng et al. [13] proposed a lightweight practical framework for fecal detection and recognition. Ozsahin et al. [14] delved into the impact of feature scaling on machine learning models for diabetes diagnosis. In the same year, Ahmad et al. [15] conducted a comparative study on sequential feature selection for cardiac disease diagnosis. Jasti et al. [16] combined structured data with image data for breast cancer diagnosis. Ahsan et al. [17] systematically analyzed multiple machine learning algorithms for cardiac diagnosis and their problems. Liao et al. [18] designed a algorithm to diagnose human health through fecal images. Yang et al. [19] proposed multi-task lightweight network for health detection based on fecal images.

In order to overcome the occlusion, image repairing is a significant progress in recent years under the impetus of deep learning technology. In 2022, Suvorov et al. [20] proposed the LaMa model, which introduced a new type of repair network with a high sensory field perceptual loss function, and was able to efficiently deal with the complex image repair tasks, such as large-area deletion. In the same year, Li et al. [21] proposed the Mask-Aware Transformer (MAT) model, which realized the efficient repair of high-resolution images by fusing the advantages of the Transformer and convolutional neural network. In 2023, Jeevan et al. [22] proposed the WavePaint model based on the WaveMix fully convolutional architecture, which further improved the efficiency of image repair.

In 2024, the field of image repair has many breakthrough achievements. The BrushNet proposed by Ju et al. [23] adopted a dual branch structure, which significantly reduced the learning burden, and could flexibly embed pre-trained diffusion models. Corneanu et al. [24] proposed LatentPaint, which achieved excellent repair results by optimizing a small number of parameters and constructing a universal diffusion model. The DNNAM proposed by Chen et al. [25] Effectively solved the problem of insufficient multi-scale feature perception by utilizing some multi-scale channel attention mechanisms. The Spa-form model proposed by Huang et al. [26] utilized sparse self attention mechanism to significantly reduce computational complexity. The PowerPaint proposed by Zhuang et al. [27], as a multifunctional repair model, achieved multitasking processing through learnable task prompts, demonstrating

strong versatility and improving repair performance.

Machine learning is effective for analyze structured data, so we use machine learning to extract features from structured data and deep learning to extract features from unstructured data. Because SS-MAT is advanced, we improve it to better diagnose skin disease.

III. METHODS

3.1. Procedure

The flowchart of skin disease diagnosis in this paper is shown in Fig. 2.

- Step 1: SelfClean
SelfClean is used to identify the occluded images.
- Step 2: Repair
SS-MAT algorithm is used to repair the occluded samples. The SS-MAT algorithm can reasonably infer the possible shape of the occluded part based on the characteristics of surrounding normal tissues, thereby achieving precise repair.
- Step 3: Image feature extraction
Deep learning models, ResNet [28] and EfficientNet [29], are used to extract image features.
- Step 4: Data mining and cleaning
Image features and structured data are processed

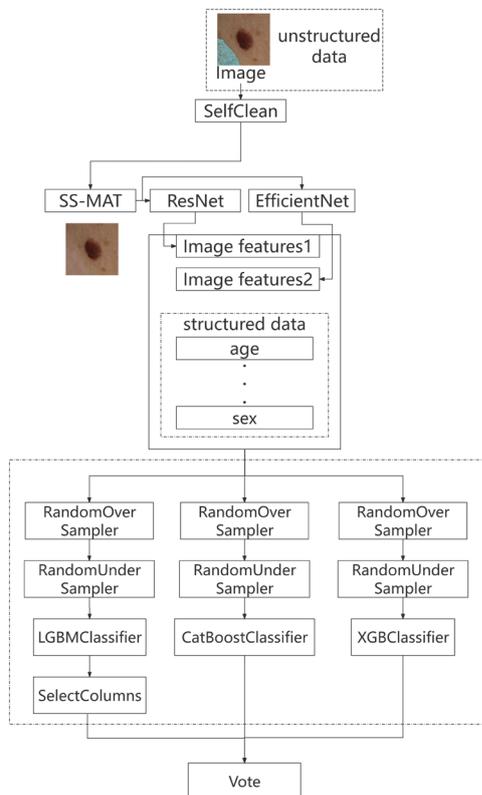


Fig. 2. Flowchart skin disease diagnosis.

through various mathematical operations to obtain new data, and then the useless and redundant features are removed, resulting in a total of 71 features.

- Step 5: Prediction
Finally, the combined features are inputted into three classifiers, namely LightGBM, CatBoost and XGBoost, which vote jointly to make the final decision.

3.2. SS-MAT

We modified and optimized SS-MAT, which is an innovative method based on mask autoencoder. Its core structure mainly consists of two main parts: encoder and decoder, as shown in Fig. 3. In the encoder stage, SS-MAT automatically masks some areas of the image. These masked areas are randomly selected, and the occlusion rate is manually set, but usually very high, e.g., 60%.

In the encoding process, only the unmasked parts of the image are processed, and SE-NET effectively improves the adaptability and sensitivity of the network to the importance of features at different channels [30], which compresses the feature map of each channel into a single value through a global average pooling operation, thus obtaining the global information between channels.

In the decoder stage, SS-MAT tries to recover the masked parts using the feature representations obtained from previous encoding that is combined with the contextual information of the image. The last step is SIREN, which has a network structure that allows learning richer and more representative features. Through the nonlinear transformation of the multi-layer sinusoidal activation function, the model can extract the features of different scales and frequencies in the video, which helps to recover the detailed information, such as textures and edges in high-resolution videos.

During the training process, SS-MAT optimizes the model parameters by minimizing the difference between

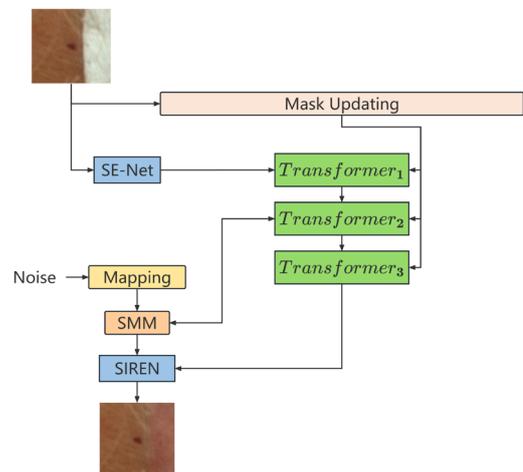


Fig. 3. SS-MAT framework.

the reconstructed image and the original image, using multiple loss functions weighted with non-saturated adversarial loss, R1 regularization, and perceptual loss to learn effective image repair patterns.

3.3. SE-Net

SE-Net in the encoder enhances the information interaction between channels in the convolutional neural network, and adaptively adjusts the response of channel features.

The SE-Net consists of three steps: Squeeze, Excitation and Scale. As shown in equation (1), where X_c denotes the c -th channel of the input feature map X , and F_{sq} denotes the squeezing operation.

$$z_c = F_{sq}(X_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X_c(i, j), c = 1, 2, \dots, C. \quad (1)$$

The excitation operation is to learn the nonlinear relationship between the channels through the two fully connected layers to generate the channel attention weights $s \in \mathbb{R}^C$. As shown in equation (2), σ is the Sigmoid function.

$$s = F_{ex}(z, W) = \sigma(W_2 ReLU(W_1 z)). \quad (2)$$

The scaling operation is to multiply the channel attention weight s with the input feature map X channel by channel to get the adjusted feature map $X' \in \mathbb{R}^{C \times H \times W}$ as:

$$X'_c = F_{scale}(X_c, s_c) = s_c \cdot X_c, c = 1, 2, \dots, C. \quad (3)$$

3.4. Attention

The traditional self-attention mechanism does not consider the mask information in the image when calculating the attention score. Let the input image features be $X \in \mathbb{R}^{N \times C}$, where N is the length of the feature sequence and C is the dimension of the features. By computing the query Q , key K and value V , the traditional self-attention is:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V. \quad (4)$$

In the mask-aware self-attention mechanism, a mask matrix $M \in \mathbb{R}^{N \times N}$ is introduced to indicate whether a pixel is in a mask region or not. In this way, the model can better distinguish the information in the masked and unmasked regions, thus capturing the global dependencies more effectively. The mask-aware self-attention is:

$$Attention_{mask}(Q, K, V) = softmax\left(\frac{QK^T + M}{\sqrt{d_k}}\right)V. \quad (5)$$

The mask M is used to identify the validity of a token, which is initialized by the input mask and automatically updated during propagation. The update rule is as follows: if at least one valid token previously existed in the window, all tokens in the window are updated to be valid after being processed by the attention mechanism. If all tokens in the window are invalid, they remain invalid after processing by the attention mechanism.

3.5. SIREN

SIREN is a neural network that uses a sinusoidal activation function to model continuous signals. The content in an image has spatial continuity, like the edges of an object, texture and so on. The sine activation function has periodicity and continuity, which can effectively capture the spatial continuity of the image. When repairing damaged images, it can fit the image details more accurately. SIREN can combine the local and global information of the image. In image repair, it is able to focus on the detailed features of the local area to be repaired, and it considers the reasonableness of the repaired content from the perspective of the overall image [31].

Let the input vector be $x \in \mathbb{R}^n$, the input of the l -th layer of SIREN be $h^{l-1} \in \mathbb{R}^{m_{l-1}}$, and the output be $h^l \in \mathbb{R}^{m_l}$. The computation of each layer of SIREN can be represented as:

$$h^l = \sin(W^l h^{l-1} + b^l). \quad (6)$$

In order to handle image information at different scales, the method adopts a multi-scale feature fusion strategy. Specifically, the model outputs feature maps at different scales in different layers, and then these feature maps are fused by a cross-scale feature fusion module. Let F_1, F_2, \dots, F_n be the feature maps of different scales, the multi-scale feature fusion module fuses these feature maps into a unified feature map F_{Fusion} through convolution and upsampling operations.

$$F_{Fusion} = \sum_{i=1}^n SIREN(Conv(UpSample(F_i))). \quad (7)$$

IV. RESULTS AND DISCUSSION

4.1. Implementation Details

The experimental environment is: Intel (R) Xeon (R) E5-2680 v4 CPU @ 2.40 GHz, 64 GB RAM, NVIDIA RTX A4000×4 GPU, Ubuntu 20.04.3 LTS 64 bit, Python 3.7, PyTorch 1.8.0 and Torchvision 0.9.0 API.

There are about 2,500 high-definition unobscured skin-cancer images in ISIC2018 and about 400,000 data in

ISIC2024. Both ISIC2018 and ISIC2024 datasets are randomly divided into training and testing sets with a ratio of 7:3.

The ISIC2018 dataset is used for the image repair task, and the image size is 256×256 , in which the occluded parts are randomly generated, and the occlusion percentage is 50%, as shown in Fig. 4. The decay coefficient of Exponential Moving Average (EMA) is 10, the learning rate is 0.001, and the batch is 4. The number of convolutional channels and the dimension of the fully connected layer for the head, body, and reconstruction modules are both set to 180. The number of blocks and the window size of the three Transformer groups, which pass through the SE-Net first and then three more Transformer groups, are $\{2, 4, 2\}$ and $\{8, 16, 8\}$, respectively. The final convolutional U-shaped network first downsamples the resolution to $1/32$. Finally the mapping network consists of 8 fully connected layers, and the outputs are implemented by SIREN and convolutional layers followed by an average pooling layer.

ISIC2024 image dataset is processed using SelfClean technology to identify the occluded images. The samples with more than half of the occluded area are discarded, and the samples with less than half of the occluded area are repaired using SS-MAT.

ResNet, EfficientNet model are used to extract the features from the image on ISIC2024 dataset with a learning rate of 0.0001, a learning rate scheduler of CosineAnnealingLR, and a number of training rounds of 50, using 5-fold cross-validation.

15 significant features are selected from the categorized data using chi-square test. Another 15 significant features are selected from all data using mutual information classification. Screening continuous data with a variance threshold of 0.05. 71 features are obtained after the aforementioned three operations.

Three classifiers, namely LightGBM, CatBoost, and XGBoost, vote jointly to make the final decision. In order to find the optimal combination of hyperparameters, hyperparameter tuning in Optuna library is performed on the classifiers. The number of training epochs for hyperparameteric tuning is 100. A soft voting strategy assigns empirical weights $[0.31, 0.45, 0.29]$ to them, respectively.

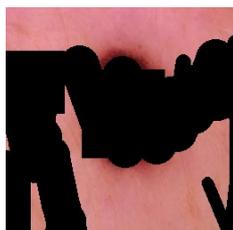


Fig. 4 Images to be repaired.

4.2. Results

4.2.1. SS-MAT Results

In medical imaging, the occlusion often interferes the observation of key lesions, bringing severe challenges to medical diagnosis. With the improved SS-MAT model, the image repair ability is more powerful, which can repair the occluded image with rich details, as shown in Fig. 5.

4.2.2. ISIC

For the binary classification of malignant examples, submissions are evaluated using the partial area under the ROC curve above an 80% true positive rate (TPR). Given the high-sensitivity requirement for cancer diagnostic systems, the evaluation metric emphasizes the area under the ROC curve with $TPR \geq 80\%$. Shown in Fig. 6, the shaded areas represent the pAUC (partial area under the ROC curve) values of two algorithms (Ca and Cb) at an arbitrary minimum TPR.

Its core idea is to compute a pAUC. The AUC value is computed within a specific false positive rate (FPR) range. Where α is set to 0.8, custom metric scaled is the pAUC value in the range of $[0, \alpha]$.



(a) Occluded image
Fig. 5. SS-MAT repair.

(b) Repaired image

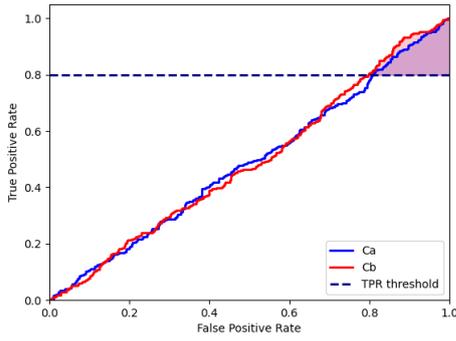


Fig. 6. pAUC values.

$$\gamma = 0.5 \times \alpha^2 + \frac{\alpha - 0.5 \times \alpha^2}{0.5}, \quad (8)$$

$$\text{custom metric} = \gamma \times (\text{custom metric scaled} - 0.5). \quad (9)$$

SelfClean scores the images with occlusions in the training data. The distribution of SelfClean is shown in Fig. 7. The smaller the irrelevance score is, the more severe the occlusion is. 0.96 means 50% occlusion.

The custom metric score is 0.1740 if only structured data are used. In order to further improve the performance, the features are thoroughly analyzed and screened, and unimportant data, which have low impacts on the predictions, are removed. After this operation, the custom metric score is increased to 0.1762. Next, in order to allow the model to learn more dimensional information, the custom metric score is improved to 0.1857 by integrating image features. Finally, in order to fully utilize the advantages of all classifiers, a soft-voting strategy is used to assign the different weights to the classifiers according to their performances, and the final custom metric score reaches 0.1935. The results are shown in Table 2.

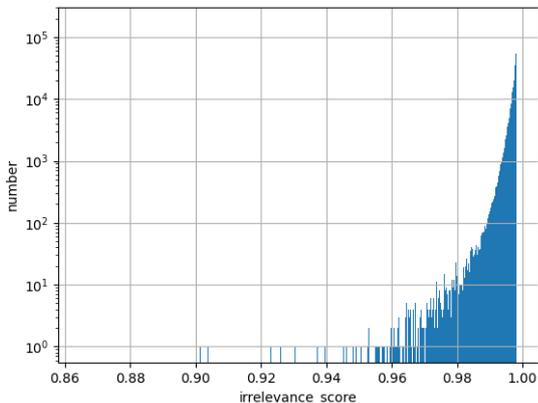


Fig. 7. Distribution of SelfClean scores.

Table 2. Prediction accuracy.

Model	Custom metric \uparrow
Only structured data	0.1740
Structured data+removal of unimportant data	0.1762
Structured data+removal of unimportant data+image features	0.1857
Structured data+remove unimportant features+image features+soft voting	0.1935

V. CONCLUSIONS AND FUTURE WORKS

An automatic diagnosis model is constructed to predict the probability that a skin cancer case is benign or malignant. ISIC2024 dataset is used, which has heterogeneous data, including structured data and unstructured data. Structured data are mined and cleaned to generate more helpful data. In unstructured data, the occluded images are selected by SelfClean and repaired by improved SS-MAT. Deep learning models are used to extract features from the repaired images. The features extracted from structured data and unstructured data are inputted to three classifiers, which vote jointly to make the final decision. In future works, we will not only focusing on occluded images but also considering other special cases such as blurred images, make the repaired images more suitable for feature extraction.

ACKNOWLEDGEMENT

This study was funded by National Natural Science Foundation of China (62466038), Jiangxi Provincial Key Laboratory of Image Processing and Pattern Recognition (2024SSY03111), Technology Innovation Guidance Program Project (Special Project of Technology Cooperation, Science and Technology Department of Jiangxi Province) (20212BDH81003), Open Foundation of Jiangxi Provincial Key Laboratory of Image Processing and Pattern Recognition (ET202404437), and Innovation Foundation for Postgraduate Students of Nanchang Hangkong University (YC2023-S746).

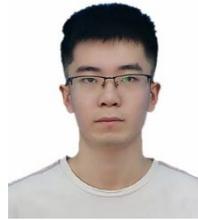
REFERENCES

- [1] A. Poniewierska-Baran, A. Poniewierska-Baran, Ł. Zadroga, E. Danilyan, P. Małkowska, and P. Niedźwiedzka-Rystwej, et al., "MicroRNA as a diagnostic tool, therapeutic target and potential biomarker in cutaneous malignant melanoma detection narrative

- review," *International Journal of Molecular Sciences*, no. 6, p. 5386, 2023.
- [2] Y. W. Lee and B. G. Kim, "Attention-based scale sequence network for small object detection," *Heliyon*, vol. 10, no. 12, 2024.
- [3] Y. J. Choi and B. G. Kim, "Hirn: Hierarchical recurrent neural network for video super-resolution (VSR) using two-stage feature evolution," *Applied Soft Computing Journal*, vol. 143, no. 110422, 2023.
- [4] K. Park and B. G. Kim, "ssFPN: Scale sequence (S2) feature-based feature pyramid network for object detection," *Sensors*, 2023.
- [5] W. Gouda, N. U. Sama, G. Al-Waakid, M. Humayun, and N. Z. Jhanjhi, "Detection of skin cancer based on skin lesion images using deep learning," *Healthcare*, vol. 10, no. 7, p. 1183, 2022.
- [6] N. Kurtansky, V. Rotemberg, M. Gillis, K. Kose, W. Reade, and A. Chow, "Isic 2024-skin cancer detection with 3D-TBP," *Slice*, 2024. <https://doi.org/10.34970/2024-slice-3d>
- [7] F. Gröger, S. Lionetti, P. Gottfrois, A. Gonzalez-Jimenez, L. Amruthalingam, and M. Groh, et al., "Intrinsic self-supervision for data quality audits," *Advances in Neural Information Processing Systems*, vol. 37, pp. 92273-92316, 2024.
- [8] N. A. Mahoto, A. Shaikh, A. Sulaiman, M. S. Al Reshan, A. Rajab, and K. Rajab, "A machine learning based data modeling for medical diagnosis," *Biomedical Signal Processing and Control*, vol. 81, p. 104481, 2023.
- [9] K. A. Bhavsar, J. Singla, Y. D. Al-Otaibi, O. Y. Song, Y. B. Zikria, and A. K. Bashir, "Medical diagnosis using machine learning: A statistical review," *Computers, Materials and Continua*, vol. 67, no. 1, pp. 107-125, 2021.
- [10] A. M. Rahmani, E. Yousefpoor, M. S. Yousefpoor, Z. Mehmood, A. Haider, and M. Hosseinzadeh, et al., "Machine learning (ml) in medicine: review, applications, and challenges," *Mathematics*, vol. 9, no. 22, p. 2970, 2021.
- [11] A. Choudhury, and D. Gupta, "A survey on medical diagnosis of diabetes using machine learning techniques," *Recent Developments in Machine Learning and Data Analytics*, pp. 1-10, Dec. 2019.
- [12] Z. Yang, L. Leng, and B. G. Kim, "StoolNET for color classification of stool medical images," *Electronics*, vol. 8, no. 12, p. 1464, 2019.
- [13] L. Leng, Z. Yang, C. Kim, and Y. Zhang, "A light-weight practical framework for feces detection and trait recognition", *Sensors*, vol. 20, no. 9, pp. 2644, 2020.
- [14] D. U. Ozsahin, M. T. Mustapha, A. S. Mubarak, Z. S. Ameen, and B. Uzun, "Impact of feature scaling on machine learning models for the diagnosis of diabetes," *IEEE*, 2022.
- [15] G. N. Ahmad, S. Ullah, A. Algethami, H. Fatima, and S. M. H. Akhter, "Comparative study of optimum medical diagnosis of human heart disease using machine learning technique with and without sequential feature selection", *IEEE Access*, vol. 10, pp. 23808-23828, 2022.
- [16] V. D. P. Jasti, A. S. Zamani, K. Arumugam, M. Naved, H. Pallathadka, and F. Sammy, et al., "Computational technique based on machine learning and image processing for medical image analysis of breast cancer diagnosis," *Security and Communication Networks*, vol. 2022, no. 1, p. 1918379, 2022.
- [17] M. M. Ahsan and Z. Siddique, "Machine learning-based heart disease diagnosis: A systematic literature review", *Artificial Intelligence in Medicine*, vol. 128, pp. 102289, Jul. 2022.
- [18] F. Liao, J. Wan, L. Leng, and C. Kim, "E-health self-help diagnosis from feces images in real scenes," *Electronics*, vol. 12, no. 2, p. 344, 2023.
- [19] Z. Yang, L. Leng, M. Li, and J. Chu, "A computer-aid multi-task light-weight network for macroscopic feces diagnosis," *Multimedia Tools and Applications*, vol. 81, no. 11, pp. 15671-15686, 2022.
- [20] R. Suvorov, E. Logacheva, A. Mashikhin, A. Remizova, A. Ashukha, and A. Silvestrov, et al., "Resolution-robust large mask inpainting with fourier convolutions," in *Conference on Applications of Computer Vision*, 2022, pp. 1-10.
- [21] W. Li, Z. Lin, K. Zhou, L. Qi, Y. Wang, and J. Jia, "Mat: Mask-aware transformer for large hole image inpainting", in *Conference on Computer Vision and Pattern Recognition*, 2022.
- [22] P. Jeevan, D. S. Kumar, and A. Sethi, "Wavepaint: Resource-efficient token-mixer for self-supervised inpainting," *arXiv Preprint arXiv:2307.00407*, 2023.
- [23] X. Ju, X. Liu, X. Wang, Y. Bian, Y. Shan, and Q. Xu, "Brushnet: A plug-and-play image inpainting model with decomposed dual-branch diffusion", *European Conference on Computer Vision*, 2024, pp. 1-15.
- [24] C. Corneanu, R. Gadde, and A. M. Martinez, "Latentpaint: Image inpainting in latent space with diffusion models", in *Conference on Applications of Computer Vision*, Jan. 2024, pp. 1-10.
- [25] Y. Chen, R. Wang, K. Yang, and K. Zou, "DNNAM: Image inpainting algorithm via deep neural networks and attention mechanism," *Applied Soft Computing*, 2024.
- [26] W. Huang, Y. Deng, S. Hui, Y. Wu, S. Zhou, and J. Wang, "Sparse self-attention transformer for image

- inpainting," *Pattern Recognition*, vol. 145, p. 109897, 2024.
- [27] J. Zhuang, Y. Zeng, W. Liu, C. Yuan, and K. Chen, "A task is worth one word: Learning with task prompts for high-quality versatile image inpainting," in *European Conference on Computer Vision*, Cham, 2024.
- [28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770-778.
- [29] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning*, vol. 97, 2019, pp. 6105-6114.
- [30] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Conference on Computer Vision and Pattern Recognition*, 2018.
- [31] Z. Chen, Y. Chen, J. Liu, X. Xu, V. Goel, and Z. Wang, et al., "Videoir: Learning video implicit neural representation for continuous space-time super-resolution," *Conference on Computer Vision and Pattern Recognition*, 2022.

AUTHORS



Zhecheng Wu is pursuing his master degree in School of Software, Nanchang Hangkong University.

His research interests include medical image segmentation and automatic medical diagnosis.



Lu Leng received his Ph.D degree from Southwest Jiaotong University, Chengdu, P. R. China, in 2012. He performed his postdoctoral research at Yonsei University, Seoul, South Korea, and Nanjing University of Aeronautics and Astronautics, Nanjing, P. R. China. He was a visiting scholar at West Virginia University, USA, and Yonsei University, South Korea.

Currently, he is a full professor, the dean of Institute of Computer Vision, the office director of Jiangxi Province Key Laboratory of Image Processing and Pattern Recognition at Nanchang Hangkong University.

Prof. Leng has published more than 100 international journal and conference papers, including more than 70 SCI papers and three highly cited papers. He has been granted several scholarships and funding projects, including six projects supported by National Natural Science Foundation of China (NSFC). He serves as a reviewer of more than 100 international journals and conferences. His research interests include computer vision, biometric template protection, biometric recognition, medical image processing, data hiding, etc.