技術論文

OSCMamba: Omni-directional Selective Scan Convolution Mamba for Medical Image Classification

Shruti Phutke Amit Shakya Chetan Gupta Rupesh Kumar Lalit Sharma

This paper, titled "OSCMamba: Omni-directional Selective Scan Convolution Mamba for Medical Image Classification," was presented at CVIP-2024 (9th International Conference on Computer Vision & Image Processing), held at IIITDM Kancheepuram, Chennai, India, from December 19–21, 2024.

Abstract

The advancement of various learning approaches has a great impact in computer vision applications specifically for medical image analysis. Being the most important task, classification accuracy of medical images has been successively improved using different methods such as Convolutional neural networks (CNNs), Transformers, etc. These models have some limitations such as the CNNs perform poorly when the feature extraction considering long-range dependency is concerned. Whereas Transformers perform well while dealing with the long-range dependencies for feature extraction, leading to the quadratic complexity. The evolution of state space models (SSMs) deals with the limitations of both the CNNs and Transformers. This has the advent of capturing the long-range dependencies and the linear complexity. Further, the scanning mechanism in the SSM provides the advantage of focusing on the required features while ignoring the rest. The existing Mamba based approach for medical image classification considers only horizontal and vertical feature scanning ignoring the diagonal information. While the omni-directional selective scan considers all of them. With this motivation, we propose omni directional selective scan-based convolution mamba (OSCMamba) approach for medical image classification. The OSCMamba approach is applied on different medical image modalities for image classification. The detailed experimental analysis with AUC and ACC on six different datasets proves the efficiency of the proposed OSCMamba based medical image classification approach. The source code is available at: https://github.com/shrutiphutke/OSCMamba

1 INTRODUCTION

Classifying the medical images into various categories depending on the clinical information present in an image has a vital role in the field of medical im- age analysis. Different medical imaging modalities such as X-Rays, Ultrasound, Optical coherence tomography (OCT), Histopathology, etc. represent the information in unique ways. Traditional diagnostic methods are often subjective, and experiments suggest that the rate of disagreement among pathologist for diagnosis is about 24%[24], highlighting the need for more consistent approaches. Computer-aided diagnosis (CAD) has emerged as a major research focus, aiming to assist radiologists by providing a second opinion that enhances accuracy, consistency, and reduces image analysis time. Recent contributions in the field of computer vision and deep learning led to precise and quick CAD[12][39] of medical images.

Extensive research on Convolution neural networks (CNNs)[54][5][15] and Transformers[36][46][41] has demonstrated significant performance gain in CAD. Detecting abnormalities in the medical images relies on the global visualization of distinct features from each organ. CNNs excel at extracting local features, but they have difficulty capturing long-range dependencies needed for global feature learning. The self-attention mechanism of Transformers deals with capturing the long-range dependencies in turn providing the global feature extractions. Despite of providing high classification accuracy, Transformers comes with high computational complexity due to its quadratic self-attention mechanism.

Due to the limitations of CNNs and Transformers, state space models (SSMs)[29] have attracted research interest for their capability to capture long-range dependencies using a linear state space framework. The conventional SSMs[19][20] have limited performance due to their limited ability of focusing or ignoring specific input features. To overcome this limitation, the advanced SSMs like Mamba[18] allows to attentively focus on relevant information via selective scanning mechanism and the hardware-aware algorithm allows linear data processing. The Mamba models are consistently used in various applications such as Language Modelling[38], Speech separation[27], Classification[58], etc. Further, the Mamba have also shown great achievement in different computer vision applications like image restoration[57], image super-resolution[7], image deblurring[16], etc.

In medical image classification, efficiently identifying and considering relevant features is crucial for achieving high classification accuracy. The existing Mamba based approach for medical image classification[53] effectively over- comes the limitations of CNNs and Transformers while ignoring the efficient feature scanning approach. With this motivation, in this work, we proposed a omni-directional selective scanning convolution block (OSCB) based approach for medical image classification. The contributions of the proposed work are:

- We propose a novel Mamba based Medical Image classification approach by exploiting the efficient feature scanning mechanism.
- The omni-directional selective scanning-based convolution block is proposed for efficient feature extraction.
- The extensive comparison of proposed method is carried out on the datasets with six different medical imaging modalities.

The consistent improvement of area under curve (AUC) and accuracy (ACC) as compared to existing state-of-the-art medical image classification methods shows the effectiveness of the proposed OSCB approach.

2 RELATED WORKS

This section gives a brief overview of the related works in the field of medical image classification and the Mamba approaches for computer vision applications.

2-1. Medical Image Classification

Classifying medical images is an essential task for diagnosing a patient’s health condition. The use of computer-aided diagnosis for medical images helps doctors analyze them more efficiently and effectively. Earlier researchers used machine learning methods for medical image classification. In this, the features of the input image are extracted with conventional computer vision approach (hand- crafted features) which are then used to train the classifier such as Support vector machine (SVM)[49], K-nearest neighbour (KNN)[35], etc. The widely used SVM classifier is applied across various medical imaging modalities, including retinopathy[40], functional magnetic resonance imaging (fMRI)[49], ultrasound[44], and others. The KNN classifier is proposed for computed tomography (CT) image classification in[35]. Similarly, Iwahori *et al.*[25] proposed K-means clustering approach for endoscopic image classification. The manual feature ex- traction followed by classification approach is time-consuming and may miss important features in the images, potentially resulting in less accurate or less reliable classification outcomes. In comparison, convolutional neural networks (CNNs) automatically learn and extract hierarchical features from raw image data, allowing them to identify intricate patterns and details more efficiently. This capability often leads to improved classification accuracy and better generalization[33].

CNN in Medical Image Classification In[42], authors proposed a CNN based approach for diagnosing the diabetic retinopathy (DR) from fundus im- ages. Similarly, Zhang *et al.*[54] proposed a CNN based approach and compared existing CNN approaches for Pneumonia detection from chest X-Ray. Further, in[5][15] the authors proposed a CNN based approach with minimal number of hidden layers for skin cancer classification. In[11], authors proposed the integration of CNN models with illumination normalization techniques to achieve higher classification accuracy. Balasubramaniam *et al.*[3] proposed a modified corrected ReLu activation based LeNet approach for breast cancer diagnosis. Several transfer learning-based methods have been proposed for medical image classification to overcome the challenge of limited datasets[45]. In[43][26] authors proposed a transfer learning on VGG, AlexNet, DenseNet201, ResNet18, SqueezeNet, etc. networks for pneumonia detection. Further, some works[22], [8], and [30] utilized the transfer learning approach for Ultrasound and retinopathy image classification respectively. Though CNN approaches achieved improved classification accuracy as compared to conventional machine learning approaches, they have limited performance due to their localized feature processing. This leads to limited ability of capturing the long-range dependency with respect to the input. The Transformers perform well in capturing the long-range dependencies which further help in efficient classification of medical images.

Transformers in Medical Image Classification Medical imaging modalities have organ specific representation that needs to be processed with highly efficient feature representation[46]. The well known fact of the Transformers of being able to efficiently capture the input dependent feature representation makes them a suitable choice over CNNs for medical image classification. Matsoukas *et al.*[37] provided the analysis of utilizing the Transformers over CNN for medical image classification task. Multiple approaches are proposed for retina disease classification[48], tumor classification[10], etc. Sun *et al.*[48] proposed a encoder with pixel relation and decoder with lesion-aware transformer for diabetic retinopathy grading. In[10], the authors proposed a hybrid CNN and Transformer based approach for multi-modal medical image classification. Gheflati *et al.*[17] proposed a similar hybrid approach with pre-trained model for breast cancer detection using ultrasound images. These approaches utilized the existing vision transformer ViT[13] model as a plug-and-play module for classification task. Further, Omid et al.[36] proposed a hybrid CNN Transformer approach with modified Transformer layer using an efficient convolution operation for medical image classification. These Transformer based approaches provide efficient classification accuracy but with increased computational complexity.

2-2. Mamba in Computer Vision

Considering the fact that CNNs capture only the local relationship of the features whilst the Transformers have quadratic complexity, researchers come up with the selective state space models called as Mamba[18]. The Mamba models are efficient at capturing the long-range dependencies unlike CNNs and have a linear computational complexity unlike Transformers. With the success of Mamba in Natural language processing tasks[38][27][50][58], it further achieved re-markable performance in vision applications[57][21][7][16]. Zheng *et al.*[57] proposed a U-shape Mamba approach for single image dehazing. In[21] a residual state space block is proposed with channel attention and Mamba with 2-D selective scanning approach for image restoration. Chen *et al.*[6] proposed a Mamba-in-mamba with 2-D selective scanning for small target detection. Further, Yue *et al.*[53] proposed a SS-Conv-SSM approach consisting of a 2-D selective scanning-based Mamba in parallel with the convolution path for medical image classification. Shi *et al.*[47] proposed a omni-selective scan by processing six directional information to overcome the unidirectional scanning limitation of 2-D selective scan for image restoration. Further, Zhao *et al.*[56] proposed an omni-directional selective scan-based approach for remote sensing image dense prediction.

Following the success of Mamba models in computer vision applications and with the utilization of efficient selective scanning mechanism, we propose a omni-directional selective scan-based convolution layer for medical image classification. The details of the proposed approach and experiments on existing dataset are provided in successive Sections.

3 METHODOLOGY

In this section, first we introduce the preliminaries of Mamba approach and then give details about the proposed omni-directional selective scan-based convolution block based approach for medical image classification.

3-1. State Space Models

The state space models (SSMs) are used to make the predictions of next state ($y(t)$) depending upon the input ($x(t)$) provided in current state ($h(t)$). The SSMs assume the inputs to be continuous in time and can be represented by two ordinary differential equations (ODEs) as state equation and the output equation.

$$\begin{aligned} h'(t) &= Ah(t) + Bx(t); \quad \textit{state equation} \\ y(t) &= Ch(t) + Dx(t); \quad \textit{output equation} \end{aligned} \tag{1}$$

where, A is the state transition matrix describing how the state $h(t)$ changes with influence of the input $x(t)$ through input projection matrix B. C indicates translation of the state towards output with respect to input $x(t)$ through the feed-forward matrix D. As the matrix D is considered as a skip connection between input and output, the SSMs are represented by neglecting the matrix D and represented as:

$$\begin{aligned} h'(t) &= Ah(t) + Bx(t); \quad \textit{state equation} \\ y(t) &= Ch(t); \qquad \textit{output equation} \end{aligned} \tag{2}$$

This represents the global feature dependency of SSMs since the current output is dependent on all the preceding states and the input. Unlike above equations where the input is considered as continuous in time, the deep learning approaches assume the input to be discrete in time. The S4[20] and Mamba[18] convert these equations from continuous ODEs to discrete time representation by utilizing zero-order hold approach. In order to achieve this, a time scale parameter Δ is introduced and the matrix A and B are transformed into $\bar{A}$ and $\bar{B}$ respectively as follows:

$$\begin{aligned} \bar{A} &= exp(\Delta A) \\ \bar{B} &= (\Delta A)^{-1}(exp(\Delta A) - I) \cdot \Delta B \end{aligned} \tag{3}$$

This discrete representation now allows to transform the discrete input x_k to the discrete output y_k with k as discrete time step by using following representation:

$$\begin{aligned} h_k &= \bar{A}h_{k-1} + \bar{A}x_k; \quad \textit{state equation} \\ y_k &= Ch_k; \qquad \textit{output equation} \end{aligned} \tag{4}$$

The SSMs can be represented with Convolution kernel ($\bar{K}$) as:

$$\begin{aligned} \bar{K} &= (C\bar{B}, C\bar{A}B, \ldots\ldots\ldots, C\bar{A}^{L-1}\bar{B}) \\ y &= x * \bar{K} \end{aligned} \tag{5}$$

where, x is input and y is output, L is length of input x. The matrix $\bar{A}$ is build with HiPPO to memorize all the hidden states in[20] called as structured state space for sequences (S4). The SSMs are efficient for modelling the input sequences but fail at filtering the irrelevant information and the ease of parallel scanning. To overcome this limitation a Selective State Space Models[58] propose a selective scan approach. Also to solve the issue of GPU utilization, hardware- aware state expansion approach is enabled in selective scan mechanism.

3-2. Proposed Architecture

In this section, we first give the pipeline of the proposed classification architecture. Further, the details of the omni-directional selective scanning convolution block followed by scanning mechanism are provided.

Architecture Overview In the proposed architecture, the input image $I \in \mathbb{R}^{H \times W \times C_{in}}$ (C_{in} = 3 for RGB image and C_{in} = 1 for gray-scale image) is firstly, converted into 4× 4 sized non-overlapping patches in Patch embedding layer to process them in Mamba block. These patches are then fed to the first omni-directional selective scan convolution block (OSCB) followed by the patch merging layer in turn generating the feature map of size $h \times w \times C$, where $h = \frac{h}{4}$, w = $\frac{w}{4}$ and $C = 96$ (see Figure 1). We call the each OSCB in Figure 1 as an encoder layer and there are four such layers with $n_l \in [2, 2, 4, 8]$ repeated OSCB where $l \in (1, 4)$ is number of layers. Further, a global average pooling followed by a fully connected layer is used to predict the output class from the input image.

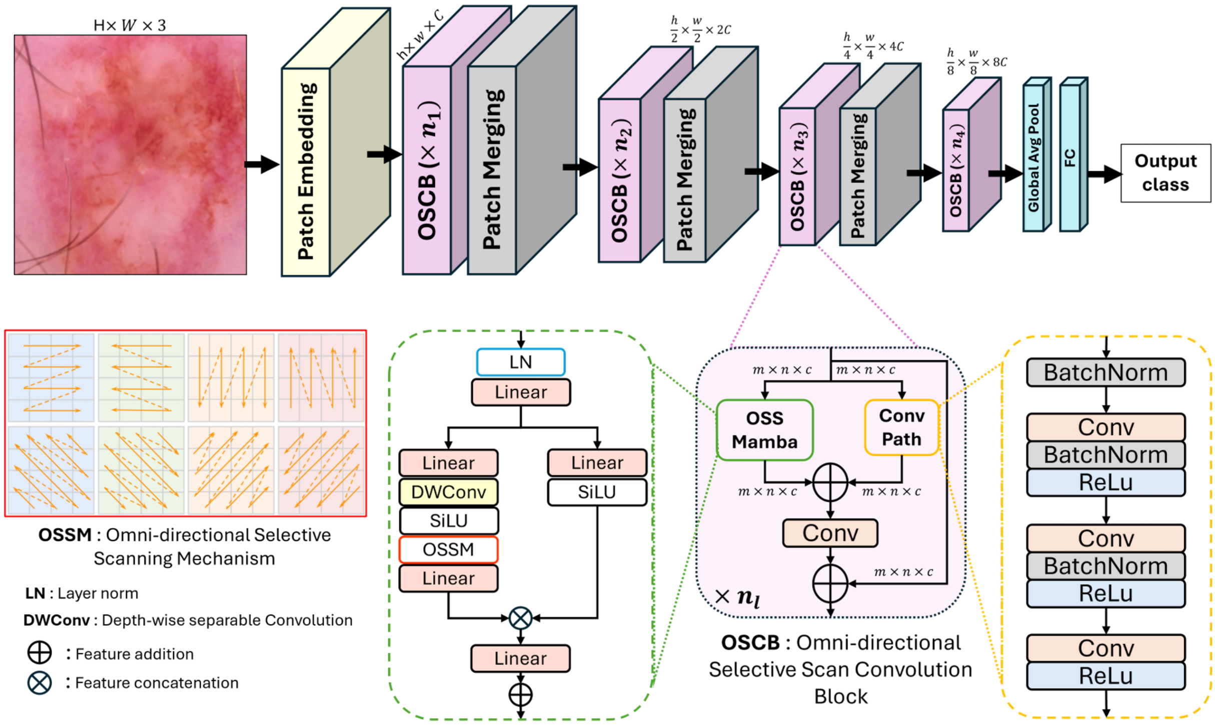


Fig. 1 Proposed Omni-directional Selective Scan Convolution mamba (OSCMamba) architecture for medical image classification

Omni-directional Selective Scanning Convolution Block Figure 1 shows the proposed omni-directive selective scan convolution block (OSCB). Unlike[53], we process the incoming features separately in omni-directional selective scan path and the convolution path. This helps the network to learn the global and local feature dependencies separately. These global and local features are then added together followed by a convolution layer. Further, a residual connection is provided to preserve the feature information (see OSCB in Figure 1). The *convolution path* (see Conv Path in OSCB of Figure 1) in OSCB processes the image with **Convolution** → **Batch Normalization** → **ReLu** to capture the local feature dependency. Further, in the *omni-directional selective scanning (OSS) Mamba* path, the incoming features are normalized and projected linearly thorough linear projection layer.

The OSS Mamba block is integration of the Gated MLP where the input features are first split into equal parts along channel dimension[55]. One path processes the features with OSS mechanism (OSSM) and the other projects the linearly convoluted features via an activation function (see the left and right split in *OSS Mamba* block of OSCB in Figure 1). In the first branch, the features are linearly projected, and a depth-wise separable convolution is applied on the features this helps in reducing the number of parameters followed by an activation function. These processed features are then forwarded to the OSSM block where these features are processed in omni-directional selective scanning mechanism in order to capture the global dependency from all the directions. These global features are then linearly projected and multiplied with the activated features from the other path (see *OSS Mamba* block in Figure 1).

Omni-directional Selective Scan Mechanism The existing selective scanning approaches such as Bi-directional, Cross-Scan, Continuous 2D, etc. consider the feature representation only in vertical and horizontal direction[55]. The Bi-directional scanning approach considers only horizontal direction for the forward and reverse scanning (similar to first two columns of the first row of Omni-directional Selective Scanning mechanism in Figure 1). Whereas the cross-scan and continuous 2D scan considers the horizontal and vertical scanning mechanism for forward and reverse approach (similar to the first row of Omni-directional Selective Scanning mechanism in Figure 1). Unlike these scanning approaches, the omni-directional selective scanning mechanism (OSSM) considers the horizontal, vertical, diagonal, and off-diagonal scanning for both forward and backward scan (see OSSM in Figure 1). These scanned features in eight directions are then processed in state space model (S6) separately[56]. Further the processed directional features are combined achieving the global information. This approach allows the efficient spatial directional feature learning.

4 EXPERIMENTS AND RESULT DISCUSSION

In this section, we provide the details of the dataset utilized for experimental analysis, the implementation details, evaluation metrics, and discussion on result analysis.

4-1. Dataset

For experimental analysis, we considered the MedMNIST[52] dataset which con- sists of set of multiple medical image modality datasets. It covers different medical imaging modalities such as Ultrasound, X-Ray, Optical coherence tomography (OCT), Dermatology, Microscope, fundus camera, etc. Due to wide variety of dataset with different classification categories such as multi-class, multi-label, binary class, etc. MedMNIST is considered as the benchmark dataset for medical image classification. Table 1 shows the training, validation and testing splits of six different datasets used for experimental analysis. The sample images from each dataset are provided in Figure 2 and details of each dataset are provided below:

BreastMNIST This dataset consists of 780 Breast Ultrasound images from three different categories: malignant, benign and normal with the original resolution of 500×500[2]. In[52], this dataset is again converted into binary classification task with positive class consisting of benign and normal images and negative class consisting of malignant images.

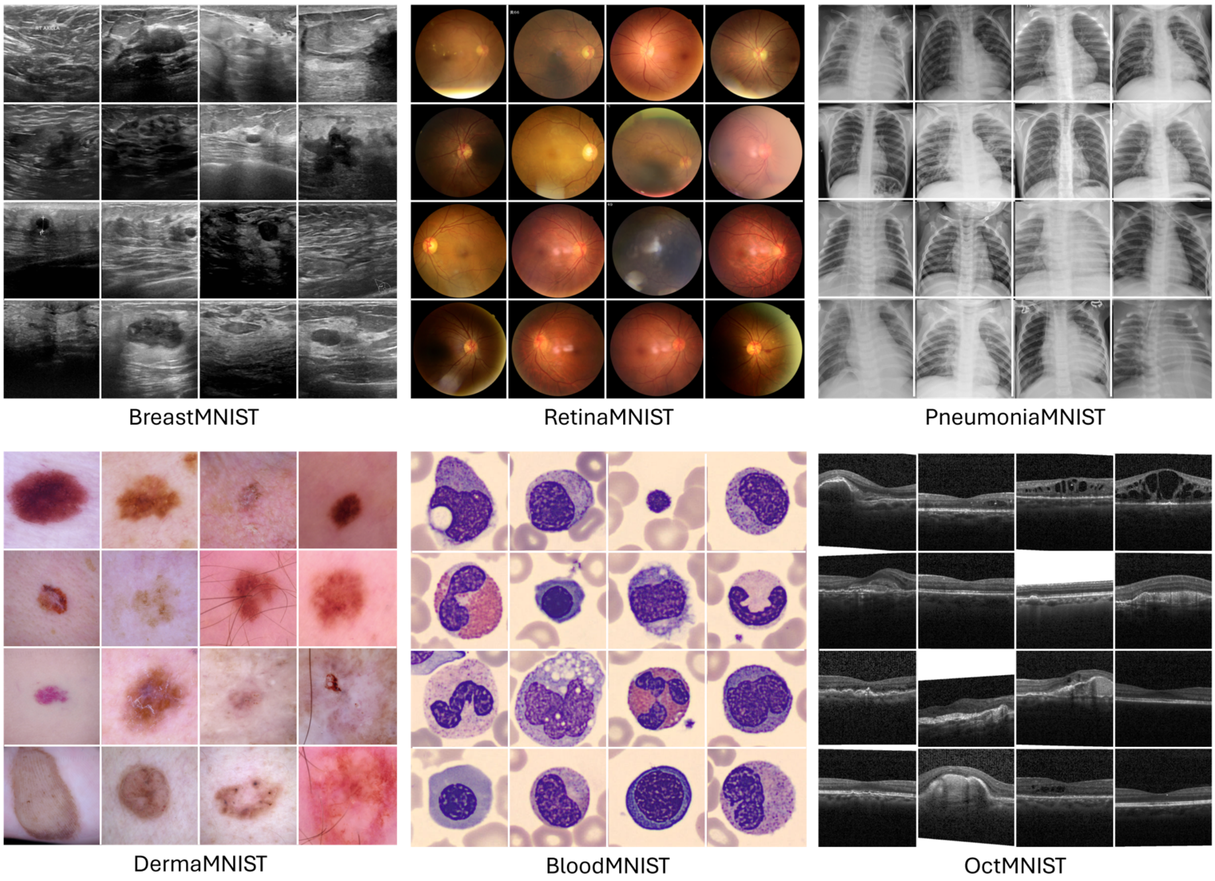
RetinaMNIST This dataset has 1600 retina fundus images with 3×1736×1824 pixel resolution from DeepDRiD dataset[34]. It has 5 different severity ratings of diabetic retinopathy. The total training dataset[34] is divided into training and validation set with a 9:1 ratio. The test set is the actual validation set from[34].

PneumoniaMNIST This is binary classification task dataset consisting of 5856 gray-scale pediatric chest X-Ray images with pixel resolution ranging in (384 − 2916)×(127 − 2713)[32][31]. The total dataset is divided into training and validation set with a 9 : 1 ratio and the test set is the actual validation set from[32][31].

DermaMNIST This dataset is a large collection of common pigmented skin lesions image from multi-source dermascope[51][9]. It has seven different categories of skin

Table 1 Overview of different datasets in MedMNIST[52]. BC: Binary-Class, OR: Ordinal Regression, MC: Multi-Class

Dataset	Data Modality	Task (#Classes/Labels)	# Train/Validation/Test
BreastMNIST	Breast Ultrasound	BC (2)	546/78/156
RetinaMNIST	Fundus Camera	OR (5)	1,080/120/400
PneumoniaMNIST	Chest X-Ray	BC (2)	4,708/524/624
DermaMNIST	Dermatoscope	MC (7)	7,007/1,003/2,005
BloodMNIST	Blood Cell Microscope	MC (8)	11,959/1,712/3,421
OCTMNIST	Retinal OCT	MC (4)	97,477/10,832/1,000



BreastMNIST RetinaMNIST PneumoniaMNIST

DermaMNIST BloodMNIST OctMNIST

Fig. 2 Sample images from different medical image classification MedMNIST dataset

lesions such as actinic keratoses and intra-epithelial carcinoma, dermatofibroma, enign keratosis-like lesions, melanoma, basal cell carcinoma, melanocytic nevi, and vascular lesions. DermaMNIST consists of 10015 dermatoscopic images with 3×600×450-pixel resolution having a 7:1:2 split ratio for training, validation and test splits.

BloodMNIST The BloodMNIST dataset consists of cellular images with 8 different categories such as basophil, erythroblast, eosinophil, immature granulocytes (metamyelocytes, myelocytes, and promyelocytes), lymphocytes, monocyte, neutrophil, and platelet[1]. This dataset is collected from the individuals that are free from any pharmacologic treatment, without any infection and hematologic or oncologic disease during the blood sample collection. It consists of 17,092 images with a training, validation and test split of 7:1:2. The images have 3×360×363 pixels resolution.

OctMNIST This dataset is derived from[31] consisting of 109309 optical coherence tomography (OCT) images for retinal disease analysis. It has 4 different categories like choroidal neovascularization, drusen, normal, diabetic macular edema. The image resolution varies in the range of (384—1536)×(277—512). The training and validation split for OctMNIST is taken from original training dataset[31] with a 9:1 split ratio and the testing split is a validation set from original dataset.

4-2. Implementation Details and Evaluation Metrics

The pre-processing on the images from all the dataset is carried out similar to MedMNIST[52]. To train the proposed classification network, we resized the images into 224×224 resolution and set the batch size = 64. The network is trained for 200 epoch using the early stopping criteria evaluated on validation accuracy with a patience of 50. The network parameters are optimized using Cross-Entropy loss and the Stochastic Gradient Descent (SGD) optimizer, with a learning rate set to 0.001. The training is carried out on NVIDIA A100 GPU. For evaluation of the performance of proposed network with state-of-the-art medical image classification methods, we considered the Area under the receiver operating characteristic (ROC) curve (AUC) and Accuracy (ACC) as evaluation metric similar to[52].

4-3. Result Analysis

The comparison of the proposed approach is done with existing state of the art medical image classification methods in terms of AUC and ACC as provided in Table 2, 3. We evaluated the proposed approach on six different medical imaging modalities dataset. As seen from the Table 2, 3 we achieve significant improvement in ACC on five among six medical image modality datasets. Where as there is significant improvement in AUC on four medical image modality datasets. Though accuracy (ACC) is more sensitive to class inconsistency

Table 2 Evaluation of the proposed method and existing state-of-the-art approaches for medical image classification. Note: **Bold** and underline shows the **best** and second best values respectively

Dataset	OCTMNIST		DermaMNIST		RetinaMNIST	
Metric	ACC	AUC	ACC	AUC	ACC	AUC
ResNet18[23]	0.943	0.743	0.917	0.735	0.717	0.524
ResNet18[23]	0.958	0.763	0.920	0.754	0.710	0.493
ResNet50[23]	0.952	0.762	0.913	0.735	0.726	0.528
ResNet50[23]	0.958	0.776	0.912	0.731	0.716	0.511
auto-sklearn[14]	0.887	0.601	0.902	0.719	0.690	0.515
AutoKeras[28]	0.955	0.763	0.915	0.749	0.719	0.503
Google AutoML[4]	0.963	0.771	0.914	0.768	0.750	0.531
MedVit-T[36]	0.961	0.767	0.914	0.768	0.752	0.534
MedVit-S[36]	0.960	0.782	0.937	0.780	**0.773**	0.561
MedVit-L[36]	0.945	0.761	0.920	0.773	0.754	0.552
MedMamba[53]	0.993	0.914	0.907	0.758	–	–
Ours	**0.995**	**0.927**	**0.948**	**0.794**	0.741	**0.573**

Table 3 Evaluation of the proposed method and existing state-of-the-art approaches for medical image classification

Dataset	PneumoniaMNIST		BloodMNIST		BreastMNIST	
Metric	AUC	ACC	AUC	ACC	AUC	ACC
ResNet18[23]	0.944	0.854	0.998	0.958	0.901	0.863
ResNet18[23]	0.956	0.864	0.998	0.963	0.891	0.833
ResNet50[23]	0.948	0.854	0.997	0.956	0.857	0.812
ResNet50[23]	0.962	0.884	0.997	0.950	0.866	0.842
auto-sklearn[14]	0.942	0.855	0.984	0.878	0.836	0.803
AutoKeras[28]	0.947	0.878	0.998	0.961	0.871	0.831
Google AutoML[4]	0.991	0.946	0.998	0.966	0.919	0.861
MedVit-T[36]	0.993	0.949	0.996	0.950	0.934	0.896
MedVit-S[36]	0.995	0.961	0.997	0.951	**0.938**	**0.897**
MedVit-L[36]	0.991	0.921	0.996	0.954	0.929	0.883
MedMamba[53]	0.965	0.912	**0.999**	0.984	0.879	0.872
Ours	**0.988**	**0.962**	**0.999**	**0.985**	0.899	0.885

than AUC, we achieve remarkable improvement in ACC as compared to state-of-the- art medical image classification methods. It is observed that, on BreastMNIST and RetinaMNIST datasets MedVit[36] achieves notable improvement. The reason behind this improvement is due to the augmentation utilized in MedVit[36], since both the BreastMNIST and RetinaMNIST have very few images in training set. Whereas in our proposed approach we have only considered horizontal flip augmentation unlike MedVit[36]. Apart from all the existing approaches, only considering a recent Mamba based classification approach[53] for comparison, we achieve the remarkable improvement on overall considered datasets in terms of AUC and ACC.

5 CONCLUSION

This work presents a novel Omni-directional selective scan convolution layer based medical image classification approach. The proposed approach is evaluated on six different medical imaging modalities such as Ultrasound, X-Ray, optical coherence tomography, histopathology, etc. The effectiveness of the proposed approach is verified by the comparison with the state-of-the-art medical image classification methods in terms of AUC and ACC. The extensive result analysis of the proposed approach verifies its effectiveness for the task of medical image classification.

REFERENCES

[1] Acevedo, A., Merino González, A., Alférez Baquero, E. S., Molina Borrás, Á., Boldú Nebot, L., Rodellar Benedé, J.: A dataset of microscopic peripheral blood cell images for development of automatic recognition systems. Data in brief 30(article 105474) (2020)

[2] Al-Dhabyani, W., Gomaa, M., Khaled, H., Fahmy, A.: Dataset of breast ultrasound images. Data in brief 28, 104863 (2020)

[3] Balasubramaniam, S., Velmurugan, Y., Jaganathan, D., Dhanasekaran, S.: A mod- ified lenet cnn for breast cancer diagnosis in ultrasound images. Diagnostics 13(17), 2746 (2023)

[4] Bisong, E., et al.: Building machine learning and deep learning models on Google cloud platform. Springer (2019)

[5] Chaturvedi, S. S., Tembhurne, J. V., Diwan, T.: A multi-class skin cancer classifica-tion using deep convolutional neural networks. Multimedia Tools and Applications 79(39), 28477–28498 (2020)

[6] Chen, T., Tan, Z., Gong, T., Chu, Q., Wu, Y., Liu, B., Ye, J., Yu, N.: Mim-istd: Mamba-in-mamba for efficient infrared small target detection. arXiv preprint arXiv:2403.02148 (2024)

[7] Cheng, C., Wang, H., Sun, H.: Activating wider areas in image super-resolution. arXiv preprint arXiv:2403.08330 (2024)

[8] Cheng, P. M., Malhi, H. S.: Transfer learning with convolutional neural networks for classification of abdominal ultrasound images. Journal of digital imaging 30, 234–243 (2017)

[9] Codella, N., Rotemberg, V., Tschandl, P., Celebi, M. E., Dusza, S., Gutman, D., Helba, B., Kalloo, A., Liopyris, K., Marchetti, M., et al.: Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). arXiv preprint arXiv:1902.03368 (2019)

[10] Dai, Y., Gao, Y., Liu, F.: Transmed: Transformers advance multi-modal medical image classification. Diagnostics 11(8), 1384 (2021)

[11] Dash, S., Parida, P., Mohanty, J. R.: Illumination robust deep convolutional neural network for medical image classification. Soft Computing pp. 1–13 (2023)

[12] Doi, K., MacMahon, H., Katsuragawa, S., Nishikawa, R. M., Jiang, Y.: Computer-aided diagnosis in radiology: potential and pitfalls. European journal of Radiology 31(2), 97–109 (1999)

[13] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)

[14] Feurer, M., Klein, A., Eggensperger, K., Springenberg, J., Blum, M., Hutter, F.: Efficient and robust automated machine learning. Advances in neural information processing systems 28 (2015)

[15] Fu'adah, Y. N., Pratiwi, N. C., Pramudito, M. A., Ibrahim, N.: Convolutional neural network (cnn) for automatic skin cancer classification system. In: IOP conference series: materials science and engineering. vol. 982, p. 012005. IOP Publishing (2020)

[16] Gao, H., Ma, B., Zhang, Y., Yang, J., Yang, J., Dang, D.: Learning enriched features via selective state spaces model for efficient image deblurring. In: ACM Multimedia (2024)

[17] Gheflati, B., Rivaz, H.: Vision transformers for classification of breast ultrasound images. In: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). pp. 480–483. IEEE (2022)

[18] Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752 (2023)

[19] Gu, A., Dao, T., Ermon, S., Rudra, A., Ré, C.: Hippo: Recurrent memory with optimal polynomial projections. Advances in neural information processing systems 33, 1474–1487 (2020)

[20] Gu, A., Goel, K., Ré, C.: Efficiently modeling long sequences with structured state spaces. arXiv preprint arXiv:2111.00396 (2021)

[21] Guo, H., Li, J., Dai, T., Ouyang, Z., Ren, X., Xia, S. T.: Mambair: A simple base- line for image restoration with state-space model. arXiv preprint arXiv:2402.15648 (2024)

[22] Gupta, S., Agrawal, S., Singh, S. K., Kumar, S.: A novel transfer learning-based model for ultrasound breast cancer image classification. In: Computational Vision and Bio-Inspired Computing: Proceedings of ICCVBIC 2022, pp. 511–523. Springer (2023)

[23] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)

[24] Hsu, W., Han, S. X., Arnold, C. W., Bui, A. A., Enzmann, D. R.: A data-driven ap- proach for quality assessment of radiologic interpretations. Journal of the American Medical Informatics Association 23(e1), e152–e156 (2016)

[25] Iwahori, Y., Hattori, A., Adachi, Y., Bhuyan, M. K., Woodham, R. J., Kasugai, K.: Automatic detection of polyp using hessian filter and hog features. Procedia computer science 60, 730–739 (2015)

[26] Jain, R., Nagrath, P., Kataria, G., Kaushik, V. S., Hemanth, D. J.: Pneumonia de-tection in chest x-ray images using convolutional neural networks and transfer learning. Measurement 165, 108046 (2020)

[27] Jiang, X., Han, C., Mesgarani, N.: Dual-path mamba: Short and long-term bidirec- tional selective structured state space models for speech separation. arXiv preprint arXiv:2403.18257 (2024)

[28] Jin, H., Song, Q., Hu, X.: Auto-keras: An efficient neural architecture search sys- tem. In: Proceedings of the 25th ACM SIGKDD international conference on knowl- edge discovery & data mining. pp. 1946–1956 (2019)

[29] Kalman, R.E.: A new approach to linear filtering and prediction problems (1960)

[30] Kandel, I., Castelli, M.: Transfer learning with convolutional neural networks for diabetic retinopathy image classification. a review. Applied Sciences 10(6), 2021 (2020)

[31] Kermany, D., Zhang, K., Goldbaum, M.: Large dataset of labeled optical coherence tomography (oct) and chest x-ray images. Mendeley Data 3(10.17632) (2018)

[32] Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C., Liang, H., Baxter, S. L., McKeown, A., Yang, G., Wu, X., Yan, F., et al.: Identifying medical diagnoses and treatable diseases by image-based deep learning. cell 172(5), 1122–1131 (2018)

[33] Li, Q., Cai, W., Wang, X., Zhou, Y., Feng, D. D., Chen, M.: Medical image clas-sification with convolutional neural network. In: 2014 13th International Confer-ence on Control Automation Robotics Vision (ICARCV). pp. 844–848 (2014). https://doi.org/10.1109/ICARCV.2014.7064414

[34] Liu, R., Wang, X., Wu, Q., Dai, L., Fang, X., Yan, T., Son, J., Tang, S., Li, J., Gao, Z., et al.: Deepdrid: Diabetic retinopathy—grading and image quality estimation challenge. Patterns 3(6) (2022)

[35] Manju, B., Meenakshy, K., Gopikakumari, R.: Prostate disease diagnosis from ct images using ga optimized smrt based texture features. Procedia Computer Science 46,

1692–1699 (2015)

[36] Manzari, O. N., Ahmadabadi, H., Kashiani, H., Shokouhi, S. B., Ayatollahi, A.: Medvit: a robust vision transformer for generalized medical image classification. Computers in Biology and Medicine 157, 106791 (2023)

[37] Matsoukas, C., Haslum, J. F., Söderberg, M., Smith, K.: Is it time to replace cnns with transformers for medical images? arXiv preprint arXiv:2108.09038 (2021)

[38] Mehta, H., Gupta, A., Cutkosky, A., Neyshabur, B.: Long range language modeling via gated state spaces. arXiv preprint arXiv:2206.13947 (2022)

[39] Melendez, J., Van Ginneken, B., Maduskar, P., Philipsen, R. H., Reither, K., Breuninger, M., Adetifa, I. M., Maane, R., Ayles, H., Sánchez, C. I.: A novel multiple-instance learning-based approach to computer-aided detection of tuberculosis on chest x-rays. IEEE transactions on medical imaging 34(1), 179–192 (2014)

[40] Niemeijer, M., Abramoff, M. D., van Ginneken, B.: Image structure clustering for image quality verification of color retina images in diabetic retinopathy screening. Medical image analysis 10(6), 888–898 (2006)

[41] Patil, P. W., Gupta, S., Rana, S., Venkatesh, S., Murala, S.: Multi-weather image restoration via domain translation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 21696–21705 (2023)

[42] Pratt, H., Coenen, F., Broadbent, D. M., Harding, S. P., Zheng, Y.: Convolutional neural networks for diabetic retinopathy. Procedia computer science 90, 200–205 (2016)

[43] Rahman, T., Chowdhury, M. E., Khandakar, A., Islam, K. R., Islam, K. F., Mahbub, Z. B., Kadir, M. A., Kashem, S.: Transfer learning with deep convolutional neural network (cnn) for pneumonia detection using chest x-ray. Applied Sciences 10(9), 3233 (2020)

[44] Rani, A., Mittal, D., et al.: Detection and classification of focal liver lesions using support vector machine classifiers. Journal of Biomedical Engineering and Medical Imaging 3(1), 21 (2016)

[45] Salehi, A. W., Khan, S., Gupta, G., Alabduallah, B. I., Almjally, A., Alsolai, H., Siddiqui, T., Mellit, A.: A study of cnn and transfer learning in medical imaging: Advantages, challenges, future scope. Sustainability 15(7), 5930 (2023)

[46] Shamshad, F., Khan, S., Zamir, S. W., Khan, M. H., Hayat, M., Khan, F. S., Fu, H.: Transformers in medical imaging: A survey. Medical Image Analysis 88, 102802 (2023)

[47] Shi, Y., Xia, B., Jin, X., Wang, X., Zhao, T., Xia, X., Xiao, X., Yang, W.: Vmambair: Visual state space model for image restoration. arXiv preprint arXiv:2403.11423 (2024)

[48] Sun, R., Li, Y., Zhang, T., Mao, Z., Wu, F., Zhang, Y.: Lesion-aware transformers for diabetic retinopathy grading. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10938–10947 (2021)

[49] Tan, L., Chen, Y., Maloney, T. C., Caré, M. M., Holland, S. K., Lu, L. J.: Combined analysis of smri and fmri imaging data provides accurate disease markers for hear- ing impairment. NeuroImage: Clinical 3, 416–428 (2013)

[50] Tang, S., Dunnmon, J. A., Liangqiong, Q., Saab, K. K., Baykaner, T., Lee-Messer, C., Rubin, D. L.: Modeling multivariate biosignals with graph neural networks and structured state space models. In: Conference on Health, Inference, and Learning. pp. 50–71. PMLR (2023)

[51] Tschandl, P., Rosendahl, C., Kittler, H.: The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific data 5(1), 1–9 (2018)

[52] Yang, J., Shi, R., Wei, D., Liu, Z., Zhao, L., Ke, B., Pfister, H., Ni, B.: Medmnist v2- a large-scale lightweight benchmark for 2d and 3d biomedical image classification. Scientific Data 10(1), 41 (2023)

[53] Yue, Y., Li, Z.: Medmamba: Vision mamba for medical image classification. arXiv preprint arXiv:2403.03849 (2024)

[54] Zhang, D., Ren, F., Li, Y., Na, L., Ma, Y.: Pneumonia detection from chest x-ray images based on convolutional neural network. Electronics 10(13), 1512 (2021)

[55] Zhang, H., Zhu, Y., Wang, D., Zhang, L., Chen, T., Wang, Z., Ye, Z.: A survey on visual mamba. Applied Sciences 14(13), 5683 (2024)

[56] Zhao, S., Chen, H., Zhang, X., Xiao, P., Bai, L., Ouyang, W.: Rs-mamba for large remote sensing image dense prediction. arXiv preprint arXiv:2404.02668 (2024)

[57] Zheng, Z., Wu, C.: U-shaped vision mamba for single

image dehazing. arXiv preprint arXiv:2402.04139 (2024)

[58] Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W., Wang, X.: Vision mamba: Efficient visual representation learning with bidirectional state space model. arXiv preprint arXiv:2401.09417 (2024)

■著者

Shruti Phutke
Emerging Technology and Innovation Lab,
Yamaha Motor Solutions India

Amit Shakya
Emerging Technology and Innovation Lab,
Yamaha Motor Solutions India

Chetan Gupta
Emerging Technology and Innovation Lab,
Yamaha Motor Solutions India

Rupesh Kumar
Emerging Technology and Innovation Lab,
Yamaha Motor Solutions India

Lalit Sharma
Emerging Technology and Innovation Lab,
Yamaha Motor Solutions India