

## **INTERNATIONAL CONFERENCE ON COMPUTER VISION** WORKSHOP ON COMPUTER VISION FOR AUTOMATED MEDICAL DIAGNOSIS







NAZARBAYEV UNIVERSITY School of Engineering and Digital Sciences

# **A Comparative Study of Vision Transformer Encoders and Few-shot Learning for Medical Image Classification**

Maxat Nurgazin and Nguyen Anh Tu Nazarbayev University, Republic of Kazakhstan

Computer vision has been significantly impacted by Vision Transformer (ViT) networks. However, most existing deep learning-based methods primarily rely on a lot of labeled data to train reliable classifiers for accurate prediction. This requirement might be impractical in the medical field.

**Problem definition:** 

Let  $D = D_1, D_2, ..., D_n$  be a collection of n medical datasets, with each dataset  $D_k$  consisting of pairs  $(x, y)_i$  representing an

This study explores the application of ViT in few-shot learning scenarios for medical image analysis, addressing the challenges posed by limited data availability. We evaluate various ViT models alongside few-shot learning algorithms, perform cross-domain experiments, and analyze the impact of data augmentation techniques.

Our findings indicate that when combined with ProtoNets, ViT architectures outperform CNN-based counterparts and achieve competitive performance against SOTA approaches on benchmark datasets.

## otivation

- $\succ$  Vision Transformers (ViTs) have emerged as an alternative to CNNs, showing impressive performance on various tasks.
- > CNNs struggle with learning long-range pixel relationships due to locality, which ViTs can handle more effectively.
- $\succ$  Medical imaging often has limited labeled data, making it difficult to train deep learning models. > Few-shot learning (FSL) is a promising approach for handling limited labeled data.

image and its label.

Result

Datasets are divided into meta-test set (D<sub>meta-test</sub>) and meta-train set (D<sub>meta-train</sub>)

Utilize abundant data in D<sub>meta-train</sub> to learn better initial weights (Reptile) or develop effective embedding space (ProtoNet & MatchingNet)

**Goal**: Improve performance on problems D<sub>meta-test</sub> with limited data (novel class data)

**Overview of the system pipeline** 



To our knowledge, ViT architectures have not been used in the field of medical image classification in few-shot learning scenarios. Therefore, given their success in other areas of computer vision, it is important to assess their performance in this area under various conditions.

### Contribution

Goa

- Investigate the efficacy of various ViT models for fewshot medical image classification.
- > Study how different few-shot learning algorithms impact the performance of ViT models.
- $\succ$  Analyze the impact of advanced data augmentation techniques on ViT models.
- > Explore the effect of a cross-domain scenario on the performance of few-shot learners.
- Framework through experiments when running on the Spark clusters.



Multi-head Attentio

- > Datasets: BreakHis (9109 microscopic images of breast tumor tissues from 82 patients with 8 classes), ISIC 2018 (10,015 dermoscopic images of skin lesions across 7 classes), and Pap Smear (917 microscopic images of cervical smears with 7 classes).
- > Experimental Settings: Pre-trained models obtained from the timm library.
  - > ProtoNet: 20 epochs, 500 episodes per epoch, SGD optimizer, learning rate of 10-5 or 10-6, cosine annealing learning rate schedule.
  - > Reptile: SGD optimizer, learning rate of 10-3 for inner optimization, learning rate of 10-1 for outer metaupdate, 1000 meta-iterations, batch size of 10 tasks, 5 and 50 adaptation steps for each task
- > Evaluation metric: Accuracy (%) as evaluation metric. 400 episodes randomly selected from novel categories in the test set. Average accuracy rate for image classification.

Algorithm	Model	I	2-way			5-way		Algorithm	Model		2-way			3-way		Algorithm	Model	FS		2-м	vay		3-w	iy
		3-shot	5-shot	10-shot	3-shot	5-shot	10-shot			3-shot	5-shot	10-shot	3-shot	5-shot	10-shot				3 sh	ot 5 sh	not 10 s	hot   3 sł	ot 5 sh	ot
ProtoNet	MViT ViT_tiny ViT_small ViT_base DeIT_base	74.64 81.03 <b>84.35</b> 83.94 72.17	76.94 83.61 <b>86.70</b> 86.02 76.53	81.50 86.52 89.72 <b>90.26</b> 81.40	60.60 67.84 72.10 72.75 57.86	64.23 71.82 76.18 <b>77.69</b> 62.38	69.23 77.68 81.45 <b>81.99</b> 69.07	Protonet	MViT ViT_tiny ViT_small ViT_base DelT_base	80.84 84.65 <b>92.40</b> 92.05 88.88	84.36 86.96 <b>94.05</b> 93.26 89.38	86.88 88.86 <b>94.90</b> 93.94 91.22	68.04 74.33 <b>86.38</b> 85.21 78.77	73.24 77.92 <b>89.09</b> 88.48 81.70	78.37 81.17 <b>90.62</b> 89.47 85.28	ProtoNet	ViT_small	Standa CutOu MixU CutMi	rt   <b>84.3</b> it   81.7 p x	5 86. 3 85.8 - -	70 89 89 89 - -	<b>.72</b> 72. .22 70. -	10 76.1 55 <b>76.2</b> -	8 3 - -
	Swin_base ResNet50 VGG16	82.49 66.62 74.11	84.17 68.65 78.17	89.12 72.81 82.11	70.75 51.43 60.68	74.67 53.83 64.58	79.92 58.34 70.84		Swin_base ResNet50 VGG16	85.42 70.49 88.75	87.56 71.75 89.34	89.78 69.61 91.76	75.73 57.74 79.04	79.88 58.48 82.53	82.46 59.60 85.63		ResNet50	Standa CutOu MixU CutMi	rt 66.6 it 65.5 p x	2 68.0 2 68.1	5 72 75 72	. <b>81 51.</b> .18 49. -	<b>53.8</b> 32 53.8 -	5 1 -
Reptile	MViT ViT_tiny ViT_small ViT_base DeIT_base Swin_base	62.80 75.80 70.30 59.30 72.80 67.30	$\begin{array}{r} 67.00 \\ 78.40 \\ 76.10 \\ 67.40 \\ \underline{79.40} \\ \overline{74.20} \end{array}$	71.80 83.50 80.40 72.70 83.00 81.10	53.00	54.47 68.67 72.13 62.27 64.73 69.00	60.33 75.13 78.53 70.53 73.60 75.93	Reptile	MViT ViT_tiny ViT_small ViT_base DeIT_base	80.60 85.60 <u>86.80</u> 77.10 84.40	80.20 88.00 <u>90.70</u> 81.50 87.30	84.30 90.10 <u>93.90</u> 88.00 92.50 87.40	72.00 75.27 77.73 70.53 76.20	73.20 82.47 82.27 78.27 82.33 81.53	78.87 86.00 87.00 88.07 86.27 87.87	Reptile	ViT_small	Standa CutOu MixU CutMi	rt 76.0 it 75.3 p <b>77.5</b> x 74.8	5         80.3           0         80.3           0         79.4           5         77.5	30     85       35     83       40     85       75     85	.55 <b>67.</b> .95 64. . <b>75</b> 66. .65 67.	50         73.1           37         69.9           20         71.3           40         72.5	5 7 3 7
	ResNet50 VGG16 MViT	71.70 64.70	72.70 72.90	76.50 78.60	47.60 56.53	51.60 62.40	54.93 70.67		ResNet50 VGG16	80.90 84.90	87.20 82.00 88.80	87.40 89.30 93.20	73.67 77.73	81.55 75.87 81.67	81.73 88.60		ResNet50	Standa CutOu MixU CutMi	rt 70.2 nt 68.7 p <b>70.7</b> x 70.1	8 75.3 3 73.0 5 74.3 0 74.0	78 78 60 76 15 78 60 77	.83         54.           .58         55.           .03         55.           .95         53	47         58.2           70         59.9           00 <b>60.6</b> 70         58.9	2 0 5 2
MatchingNat	ViT_tiny ViT_small ViT_base	76.66 78.40 79.81 73.67 72.84 67.99 72.20	79.88 81.61 <u>84.19</u> 77.11	83.41 86.34 <u>88.21</u> 81.60	63.42 65.50 <u>67.70</u> 58.23 58.66	$ \begin{array}{r} 66.62 \\ 70.00 \\ \underline{73.30} \\ 62.54 \\ 63.09 \end{array} $	71.95 76.47 <u>79.17</u> 70.47 68.34	MatchingNet	ViT_tiny ViT_small ViT_base DelT_base Swin_base	80.10 86.00 <u>90.84</u> 89.56 89.25 82.01	81.97 89.09 <u>92.56</u> 89.50 89.36 83.58	85.32 90.91 <u>94.27</u> <u>92.10</u> 91.70 86.94	67.61 77.34 <u>84.74</u> 78.24 79.43 70.34	80.92 87.02 82.53 82.39 74.38	70.35 83.93 89.23 86.33 85.16 78.27	Effect Few-s	of diffe	erent	Augn	nenta for IS	tion t	echnic	ues o ataset	n
Matchingivet	DelT_base Swin_base ResNet50 VGG16	72.84 67.99 72.20	75.60 71.69 75.94	80.12 75.66 79.90	52.93 59.76	56.53 61.63	61.60 67.98		ResNet50	76.01	74.46	77.89	59.66	61.23	64.08			551110						
Donformer	DefT_base Swin_base ResNet50 VGG16	72.84 67.99 72.20	75.60 71.69 75.94	80.12 75.66 79.90	52.93 59.76	56.53 61.63	61.60 67.98		ResNet50 VGG16	76.01 87.66	74.46 88.30	77.89 89.94	59.66 77.07	61.23 79.60	64.08 83.45	Target	s	etting		2-way			3-way	
Perform learning	ance of algorith	72.84 67.99 72.20 mode hms f	75.60 71.69 75.94 els us for IS	<sup>80,12</sup> 75.66 79.90 ing di [C 20]	fferer 18 dat	taset	61.60 67.98 ta-	Perform	nance of g algorith	mode	r4.46 88.30	<sup>77.89</sup> 89.94 ng dif	ferent	<sup>61.23</sup> 79.60 t meta aset	64.08 83.45	Target	x100 CD	etting   + PN	3-shot 74.12	2-way 5-shot 78.74	10-shot 84.11	3-shot   62.72	3-way 5-shot 68.90	
Perform learning	ance of algorith	72.84 67.99 72.20 mode hms f	75.60 71.69 75.94	<sup>80,12</sup> 75.66 79.90 ing di [C 20]	fferen 18 dat	<sup>56.53</sup> 61.63 nt me taset	61.60 67.98 ta-	Perform learnin	nance of g algorith	mode	els usin or Pap	<sup>77.89</sup> 89.94 ng dif Smea	<sup>59.66</sup> 77.07 ferent ar data	61.23 79.60 t meta aset	64.08 83.45	Target BreakHis 2	X100 CD No	etting   + PN   n CD	3-shot 74.12 80.64	2-way 5-shot 78.74 83.80	10-shot 84.11 87.62	3-shot 62.72 69.39	3-way 5-shot 68.90 <b>75.91</b>	
Perform learning Algorith	ance of algorith	72.84 67.99 72.20 mode hms f	75.60 71.69 75.94 els us for IS	80.12 75.66 79.90 ing di [C 20] 2-way 5-shot	fferer 18 dat	<sup>56.53</sup> 61.63 nt me taset	61.60 67.98 ta- 3-way 5-shot 10	Perforn learnin	nance of g algorith	$\frac{76.01}{87.66}$ mode	74.46 88.30 els usin or Pap 2-way t 5-shot	77.89 89.94 ng dif 5 Smea	59.66 77.07 Ferent ar data	61.23 79.60 t meta aset 3-way 5-shot	64.08 83.45 <b>A-</b>	Target BreakHis 2 Pap Smear	X100 CD No . CD No	+ PN n CD + PN n CD	3-shot 74.12 80.64 92.22 92.40	2-way 5-shot 78.74 <b>83.80</b> <b>94.12</b> 94.05	10-shot 84.11 87.62 94.85 94.90	3-shot 62.72 69.39 86.22 86.38	3-way 5-shot 68.90 <b>75.91</b> 88.82 <b>89.09</b>	
Perform learning Algorith ProtoNet	ance of algorith	72.84         67.99         72.20         mode         hms         Model         [         F_small         ssNet50	75.60 71.69 75.94 els us for IS 3-shot 84.35 66.62	80.12 75.66 79.90 ing di [C 20] 2-way 5-shot 1 86.70 68.65	10-shot         1           89.72         72.81	<sup>56.53</sup> 61.63 nt me taset <sup>3-shot</sup> 72.10 51.43	61.60 67.98 ta- 3-way 5-shot 10 76.18 8 53.83 5	Perform learnin -shot -s	ResNet50 VGG16 nance of g algorith Mod	$\frac{1000}{76.01} \\ 87.66 \\ mode \\ ms fe \\ 1000 \\ 10$	74.46 88.30 els usin or Pap 2-way t 5-shot 4 83.80 2 72.12	77.89 89.94 ng dif 5 Smea 7 10-shot 87.62 73.31	59.66 77.07 Ferent ar data   3-shot   69.39 55.80	61.23 79.60 t meta aset 3-way 5-shot 75.91 60.28	64.08 83.45 <b>A-</b> 10-shot 81.47 61.88	Target BreakHis 2 Pap Smear	x100 CD No CD No CD No CD	etting   + PN   n CD   + PN   n CD	3-shot 74.12 80.64 92.22 92.40	2-way 5-shot 78.74 83.80 94.12 94.05 8-to-]	10-shot 84.11 87.62 94.85 94.90 Break	3-shot   62.72   69.39   86.22   86.38 His at	3-way 5-shot 68.90 75.91 88.82 89.09	p :
Perform learning Algorith ProtoNet Matching	ance of algorith ang ViT gNet	72.84         67.99         72.20         mode         hms         Model         -         F_small         ssNet50         F_small         ssNet50	75.60 71.69 75.94 els us for IS 3-shot 84.35 66.62 78.40 67.99	80.12 75.66 79.90 ing di [C 20] 2-way 5-shot 1 86.70 68.65 81.61 71.69	10-shot   <b>89.72</b> 72.81 86.34 75.66	<sup>56.53</sup> 61.63 <b>nt me</b> <b>taset</b> 3-shot <b>72.10</b> 51.43 65.50 52.93	61.60 67.98 ta- 3-way 5-shot 10 76.18 8 53.83 5 70.00 7 56.53 6	Perform learnin Algorithm shot 41.45 8.34 ProtoNet 6.47 Matching	ResNet50 VGG16 nance of g algorith Mod ViT_sma ResNet5	76.01 87.66 mode ms fe el	74.46 88.30 21s usin pr Pap 2-way t 5-shot 4 83.80 2 72.12 3 82.09 5 76.58	77.89 89.94 ng dif 5 Smea 7 10-shot 87.62 73.31 88.33 79.14	59.66 77.07 Ferent ar data 3-shot 69.39 55.80 67.13 59.53	61.23 79.60 t meta aset 3-way 5-shot 75.91 60.28 72.88 62.70	64.08 83.45 <b>A-</b> 10-shot 81.47 61.88 <b>81.80</b> 66.98	Target BreakHis 2 Pap Smear Cross	x100 CD No CD No CD No	etting   + PN   n CD   + PN   n CD	3-shot 74.12 80.64 92.22 92.40	2-way 5-shot 78.74 83.80 94.12 94.05 8-to-]	10-shot 84.11 87.62 94.85 94.90 Break	3-shot   62.72   69.39   86.22   86.38 His at	3-way 5-shot 68.90 75.91 88.82 89.09	p :
Perform learning Algorith ProtoNet Matching	ance of algorith algorith ance vin algorith	72.84         67.99         72.20         mode         hms         hms         f         Model            f_small         esNet50         f_small         esNet50	75.60 71.69 75.94 els us or IS or IS 3-shot 84.35 66.62 78.40 67.99 76.05 70.28	80.12 75.66 79.90 ing di [C 20] 2-way 5-shot 86.70 68.65 81.61 71.69 80.30 75.78	10-shot         1           89.72         72.81           86.34         75.66           85.55         78.92	<sup>56.53</sup> 61.63 <b>nt me</b> <b>taset</b> <sup>3-shot</sup> <b>72.10</b> 51.43 65.50 52.93 67.50 54.47	61.60 67.98 ta- 3-way 5-shot 10 76.18 8 53.83 5 70.00 7 56.53 6 73.15 7 58.22 7	Perform learnin Algorithm -shot -shot -shot -shot -shot -shot -shot -shot -shot -shot -shot 	ResNet50 VGG16 nance of g algorith Mod ViT_sma ResNet5 Net ViT_sma	$\frac{1000}{76.01} \\ 87.66 \\ mode \\ ms fe \\ 10 \\ 11 \\ 10 \\ 11 \\ 10 \\ 10 \\ 10 \\ 1$	74.46         88.30         els usin         or Pap         2-way         t       5-shot         4       83.80         2       72.12         3       82.09         5       76.58         0       75.60	77.89 89.94 ng dif 5 Smea 7 10-shot 87.62 73.31 88.33 79.14 81.60	59.66         77.07         ferent         ar data           3-shot         69.39         55.80         67.13         59.53         54.40	61.23 79.60 t meta aset 3-way 5-shot 75.91 60.28 72.88 62.70 63.13	64.08 83.45 <b>A-</b> 10-shot 81.47 61.88 <b>81.80</b> 66.98 72.20	Target BreakHis 2 Pap Smear	x100 CD No CD No CD No	+ PN n CD + PN n CD n CD	3-shot 74.12 80.64 92.22 92.40	2-way 5-shot 78.74 83.80 94.12 94.05 8-to-]	10-shot 84.11 87.62 94.85 94.90 Break	3-shot   62.72   69.39   86.22   86.38	3-way 5-shot 68.90 <b>75.91</b> 88.82 <b>89.09</b> nd -Pa	p ;
Perform learning Algorith ProtoNet Matching Reptile	ance of algorith ans vir algorith ans vir algorith algorith ans vir algorith ans vir algorith ans vir algorith ans vir algorith	72.84         67.99         72.20         mode         hms         f         Model         [         F_small         essNet50         F_small         essNet50         F_small         essNet50         Med[20]	75.60 71.69 75.94 els us for IS 3-shot 84.35 66.62 78.40 67.99 76.05 70.28 72.75	80.12 75.66 79.90 ing di [C 20] 2-way 5-shot 1 86.70 68.65 81.61 71.69 80.30 75.78 75.62	50.00         52.93         59.76         fferer         18 dat         10-shot         89.72         72.81         86.34         75.66         85.55         78.83         81.37	56.53 61.63 <b>nt me</b> taset 3-shot 72.10 51.43 65.50 52.93 67.50 54.47 54.83	61.60 67.98 ta- 3-way 5-shot 10 76.18 8 53.83 5 70.00 7 56.53 6 73.15 7 58.22 6 59.33 6	Perform learnin Algorithm shot Algorithm shot ProtoNet 6.47 Matching 7.37 51.58 9.75	ResNet50 VGG16 nance of g algorith Mod ViT_sma ResNet5 Net ViT_sma ResNet5 ViT_sma ResNet52 ViT_sma ResNet52	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	74.46 88.30 215 USI 2-way t 5-shot 4 83.80 2 72.12 3 82.09 5 76.58 0 75.60 5 81.38 6 85 28	77.89 89.94 ng dif Smea 7 10-shot 87.62 73.31 88.33 79.14 81.60 83.88	59.66 77.07 Ferent ar data 3-shot 69.39 55.80 67.13 59.53 67.13 59.53 54.40 63.08	61.23 79.60 t meta aset 3-way 5-shot 75.91 60.28 72.88 62.70 63.13 66.42 75.04	64.08 83.45 <b>2-</b> 10-shot 81.47 61.88 <b>81.80</b> 66.98 72.20 74.08 78.93	Target BreakHis 2 Pap Smear Cross	x100 CD No CD No -domai	etting  _ + PN   n CD   + PN   n CD	3-shot 74.12 80.64 92.22 92.40	2-way 5-shot 78.74 83.80 94.12 94.05 8-to-]	10-shot 84.11 87.62 94.85 94.90 Break	3-shot   62.72   69.39   86.22   86.38	3-way 5-shot 68.90 <b>75.91</b> 88.82 <b>89.09</b>	p :
Perform learning Algorith ProtoNet Matching Reptile	Deff_base Swin_base ResNet50 VGG16 ance of algorith ance of algorith um t ViT Res gNet ViT Res ViT Res MetaM	72.84         67.99         72.20         mode         hms         f         Model         [         F_small         ssNet50         F_small         ssNet50         F_small         ssNet50         Med[0]         Med[6]	75.60 71.69 75.94 els us for IS 3-shot 84.35 66.62 78.40 67.99 76.05 70.28 72.75 81.69	80.12 75.66 79.90 ing di [C 20] 2-way 5-shot 1 86.70 68.65 81.61 71.69 80.30 75.78 75.62 83.87	10-shot         1           10-shot         1           89.72         72.81           86.34         75.66           85.55         78.83           81.37         85.14	56.53 61.63 <b>nt me</b> taset 3-shot <b>72.10</b> 51.43 65.50 52.93 67.50 54.47 54.83 66.94	61.60 67.98 ta- 3-way 5-shot 10 76.18 8 53.83 5 70.00 7 56.53 6 73.15 7 58.22 6 59.33 6 69.78 7	Perform learnin Algorithm shot 7.37 61.58 9.75 3.81	ResNet50 VGG16 nance of g algorith Mod ViT_sma ResNet5 Net ViT_sma ResNet5 ViT_sma MetaMed[20 PFEMed[0	$\begin{array}{c c}     76.01 \\     87.66 \\ \hline     mode \\     ms fe \\     el                              $	74.46 88.30 215 USI 2-way t 5-shot 4 83.80 2 72.12 3 82.09 5 76.58 0 75.60 5 81.38 6 85.28	77.89 89.94 ng dif Smea 7 10-shot 87.62 73.31 88.33 79.14 81.60 83.88 86.90	59.66 77.07 Ferent ar data   3-shot   69.39 55.80   67.13 59.53   54.40   63.08   69.21	61.23 79.60 t meta aset 3-way 5-shot 75.91 60.28 72.88 62.70 63.13 66.42 75.04	64.08 83.45 <b>2-</b> 10-shot 81.47 61.88 <b>81.80</b> 66.98 72.20 74.08 78.93	Target BreakHis 2 Pap Smear	x100 CD No CD No -domai	etting   + PN   n CD   + PN   n CD	3-shot 74.12 80.64 92.22 92.40	2-way 5-shot 78.74 83.80 94.12 94.05 8-to-]	10-shot 84.11 87.62 94.85 94.90 Break	3-shot   62.72   69.39   86.22   86.38	3-way 5-shot 68.90 75.91 88.82 89.09	

> Our methods achieve state-of-the-art performance on challenging medical datasets of few-shot medical image classification

- Zhiyong et al., Pfemed: Few- shot medical image classification using prior guided feature enhancement. Pattern Recognition, 2023.
- Rishav et al., Metamed: Few-shot medical image classification using gradient-based meta-learning. Pattern Recognition 2021.
- Yisheng et al., A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities. arXiv preprint arXiv:2205.06743, 2022