Supplementary Material

Wireless Capsule Endoscopy Multiclass Classification Using 3D Deep CNN Model

Mehrdokht Bordbar, Mohammad Sadegh Helfroush, Habibollah Danyali, Fardad Ejtehadi

Computational and Mathematical Methods in Medicine

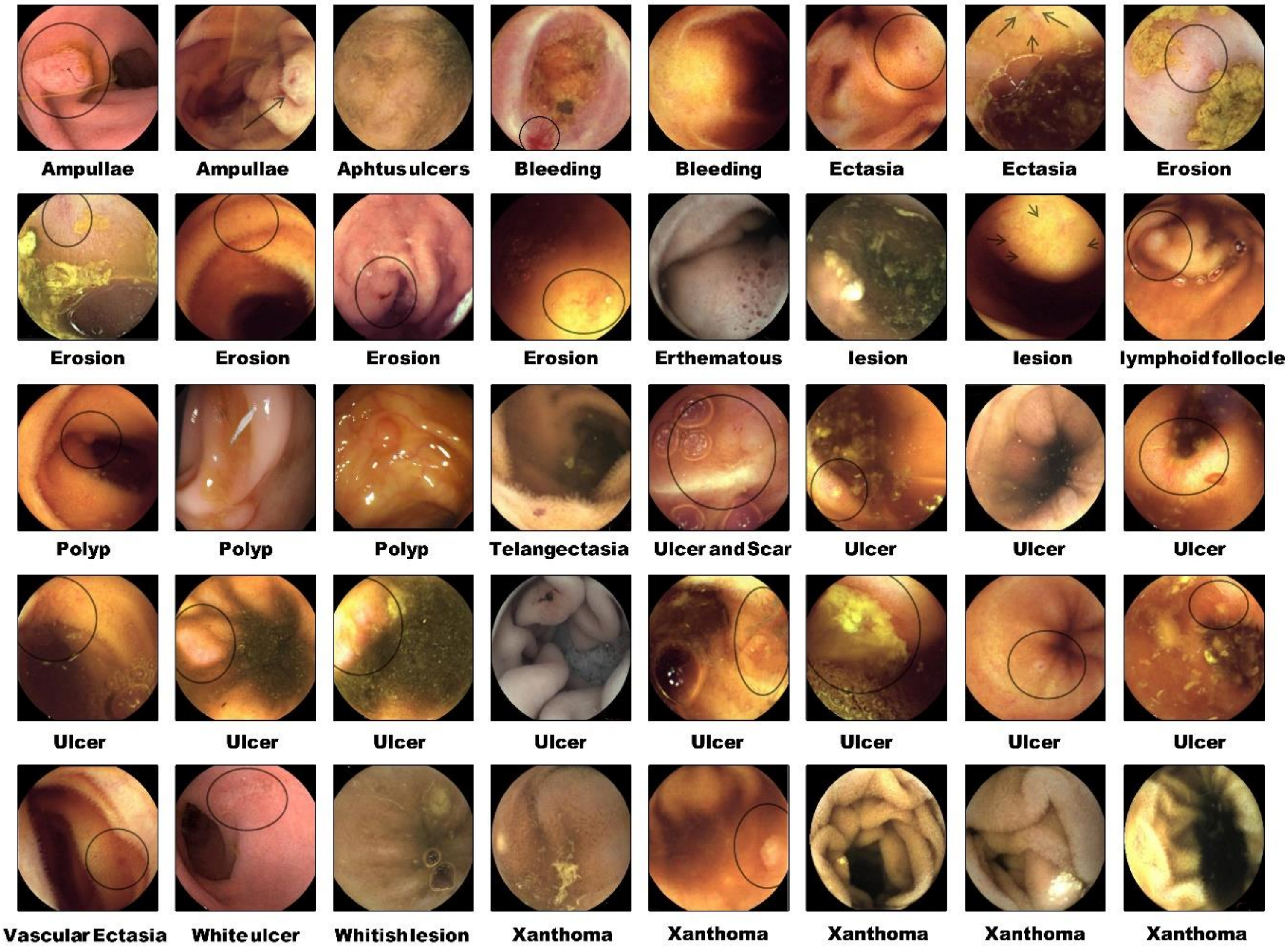

Figure S1. Samples of different types of lesions in the used dataset

Table S1 The comparisons of numerous studies in WCE video classification using deep networks

Year	Method	Task	Dataset	Validation method	Results
2016 [1]	8 Layer CNN	bleeding detection and classification	10,000 frames (Proprietary videos)	Cross validation, frame-based	F1 score: 99.5, Precision: 99.9, Recall: 99.2
2018 [2]	AlexNet	Ulcer and erosion detection and classification	21,160 frames (Proprietary videos)	Cross validation, frame-based	Accuracy: 95 Sensitivity: 95.2, Specificity: 95.7
2019 [3]	AlexNet, GooglNet	Ulcer detection and classification	1,875 frames (Proprietary videos)	Split dataset, frame-based	Accuracy: 100 Sensitivity: 100, Specificity: 100
2019 [4]	Single Shot Multi-Box Detector	angioectasia detection and classification	2,237 frames (Proprietary videos)	Split dataset, frame-based	Accuracy: 98.7 Sensitivity: 98.8 Specificity: 98.4
2019 [5]	Single Shot Multi-Box Detector	Ulcers and erosion detection and classification	1,580 frames (Proprietary videos)	Split dataset, disease-based	Accuracy: 90.8 Sensitivity: 88.2 Specificity: 90.9
2020 [6]	Inception-ResNet-v2	Significant frames classification	210,100 frames (Proprietary videos)	Cross validation, frame-based & disease-based	Accuracy: 98.3 Sensitivity: 96 Specificity: 99.5
2021 [7]	AlexNet + Triplet Network	Abnormality detection and classification	5,306 frames (Proprietary videos)	Split dataset, frame-based	Accuracy: 90.8 Sensitivity: 91.4 Specificity: 90.9
2021 [8]	MobileNet + CNN	bleeding detection and classification	1650 frames (Proprietary videos)	Cross validation, frame-based	Accuracy: 99.3 Sensitivity: 99.4 Precision: 100
2021 [9]	Inception-Resnet-V2	Multiple disease detection and classification	400,000 frames (Proprietary videos)	Split dataset, frame-based	Accuracy: 97.9 Sensitivity: 97.2 Specificity: 98.7
2021 [10]	Backbone networks + SVM (VGG19, InceptionV3, ResNet50)	bleeding detection and classification	KID Dataset 2 [11] , MICAAI 2017 [12]	Cross validation, frame-based	Accuracy: 98.2, 95.7 Sensitivity: 98.7, 95.8 Specificity: 97.3, 95.5
2022 [13]	Backbone networks + Dense layers (Efficientnet, DenseNet, Xception)	Multiple disease detection and classification	Kvasir [14]	Split dataset, frame-based	Accuracy: 94.8

Table S2. Proposed 2D-CNN characteristics

Layers arrange and their specifications		
1	Image input	224×224×3 images with "zerocenter" normalization
2	Convolution	32 3×3×3 convolutions with stride [1 1] and padding "same"
3	Batch normalization	Batch normalization with 32 channels
4	ReLU	ReLU activation function
5	Max pooling	2×2 max pooling with stride [2 2] and "zero" padding
6	Dropout	Dropout with 0.4 rate
7	Convolution	64 3×3×32 convolutions with stride [1 1] and padding "same"
8	Batch normalization	Batch normalization with 64 channels
9	ReLU	ReLU activation function
10	Max pooling	2×2 max pooling with stride [2 2] and "zero" padding
11	Dropout	Dropout with 0.4 rate
12	Convolution	128 3×3×64 convolutions with stride [1 1] and padding "same"
13	Batch normalization	Batch normalization with 128 channels
14	ReLU	ReLU activation function
15	Max pooling	2×2 max pooling with stride [2 2] and "zero" padding
16	Dropout	Dropout with 0.4 rate
17	Convolution	256 3×3×128 convolutions with stride [1 1] and padding "same"
18	Batch normalization	Batch normalization with 256 channels
19	ReLU	ReLU activation function
20	Max pooling	2×2 max pooling with stride [2 2] and "zero" padding
21	Dropout	Dropout with 0.4 rate
22	Convolution	512 3×3×256 convolutions with stride [1 1] and padding "same"
23	Batch normalization	Batch normalization with 512 channels
24	ReLU	ReLU activation function
25	Max pooling	2×2 max pooling with stride [2 2] and "zero" padding
26	Dropout	Dropout with 0.4 rate
27	Fully connected	1 fully connected layer after the flatten filter
28	Dropout	Dropout with 0.4 rate
29	Softmax	Softmax layer with 3 output classes
Training options		
	Max epochs	150
	Batch size	32
	Learning rate	0.0001
	Decay rate	0.0001
	Solver for training network	Adam optimizer
	Loss function	Cross-entropy

Table S3. Proposed 3D-CNN characteristics

Layers arrange and their specifications		
1	Image input	224×224×3×16 images with "zerocenter" normalization
2	Convolution	32 3×3×3×3 convolutions with stride [1 1 1] and padding "same"
3	Batch normalization	Batch normalization with 32 channels
4	ReLU	ReLU activation function
5	Max pooling	2×2×1 max pooling with stride [2 2 2] and "zero" padding
6	Dropout	Dropout with 0.4 rate
7	Convolution	64 3×3×3×32 convolutions with stride [1 1 1] and padding "same"
8	Batch normalization	Batch normalization with 64 channels
9	ReLU	ReLU activation function
10	Max pooling	2×2×2 max pooling with stride [2 2 2] and "zero" padding
11	Dropout	Dropout with 0.4 rate
12	Convolution	128 3×3×3×64 convolutions with stride [1 1 1] and padding "same"
13	Batch normalization	Batch normalization with 128 channels
14	ReLU	ReLU activation function
15	Max pooling	2×2×2 max pooling with stride [2 2 2] and "zero" padding
16	Dropout	Dropout with 0.4 rate
17	Convolution	256 3×3×3×128 convolutions with stride [1 1 1] and padding "same"
18	Batch normalization	Batch normalization with 256 channels
19	ReLU	ReLU activation function
20	Max pooling	2×2×2 max pooling with stride [2 2 2] and "zero" padding
21	Dropout	Dropout with 0.4 rate
22	Convolution	512 3×3×3×256 convolutions with stride [1 1 1] and padding "same"
23	Batch normalization	Batch normalization with 512 channels
24	ReLU	ReLU activation function
25	Max pooling	2×2×2 max pooling with stride [2 2 2] and "zero" padding
26	Dropout	Dropout with 0.4 rate
27	Fully connected	1 fully connected layer after the flatten filter
28	Dropout	Dropout with 0.4 rate
29	Softmax	Softmax layer with 3 output classes
Training options		
	Max epochs	150
	Batch size	32
	Learning rate	0.0001
	Decay rate	0.0001
	Solver for training network	Adam optimizer
	Loss function	Cross-entropy

References:

[1] X. Jia and M. Q.-H. Meng, "A deep convolutional neural network for bleeding detection in wireless capsule endoscopy images," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 639-642, 2016.

[2] S. Fan, L. Xu, Y. Fan, K. Wei, and L. Li, "Computer-aided detection of small intestinal ulcer and erosion in wireless capsule endoscopy images," *Physics in Medicine & Biology,* vol. 63, p. 165001, 2018.

[3] H. Alaskar, A. Hussain, N. Al-Aseem, P. Liatsis, and D. Al-Jumeily, "Application of convolutional neural networks for automated ulcer detection in wireless capsule endoscopy images," *Sensors,* vol. 19, p. 1265, 2019.

[4] A. Tsuboi and S. Oka, "Artificial intelligence using a convolutional neural network for automatic detection of small-bowel angioectasia in capsule endoscopy images," *Digestive Endoscopy,* vol. 32(3), pp. 382-390, 2020.

[5] T. Aoki, A. Yamada, K. Aoyama, H. Saito, A. Tsuboi, A. Nakada, R. Niikura, M. Fujishiro, S. Oka, and S. Ishihara, "Automatic detection of erosions and ulcerations in wireless capsule endoscopy images based on a deep convolutional neural network," *Gastrointestinal endoscopy,* vol. 89, pp. 357-363. e2, 2019.

[6] J. Park, Y. Hwang, J. H. Nam, D. J. Oh, K. B. Kim, H. J. Song, S. H. Kim, S. H. Kang, M. K. Jung, and Y. Jeong Lim, "Artificial intelligence that determines the clinical significance of capsule endoscopy images can increase the efficiency of reading," *PLoS One,* vol. 15, p. e0241474, 2020.

[7] S. Adewole, P. Fernandes, J. Jablonski, A. Copland, M. Porter, S. Syed, and D. Brown, "Lesion2Vec: Deep Meta Learning for Few-Shot Lesion Recognition in Capsule Endoscopy Video," in *Proceedings of the Future Technologies Conference (FTC) 2021, Volume 2*, Cham, pp. 762-774, 2022.

[8] F. Rustam, M. A. Siddique, H. U. R. Siddiqui, S. Ullah, A. Mehmood, I. Ashraf, and G. S. Choi, "Wireless Capsule Endoscopy Bleeding Images Classification Using CNN Based Model," *IEEE Access,* vol. 9, pp. 33675-33688, 2021.

[9] S. H. Kim, Y. Hwang, D. J. Oh, J. H. Nam, K. B. Kim, J. Park, H. J. Song, and Y. J. Lim, "Efficacy of a comprehensive binary classification model using a deep convolutional neural network for wireless capsule endoscopy," *Scientific Reports,* vol. 11, p. 17479, 2021.

[10] A. Caroppo, A. Leone, and P. Siciliano, "Deep transfer learning approaches for bleeding detection in endoscopy images," *Computerized Medical Imaging and Graphics,* vol. 88, p. 101852, 2021.

[11] A. Koulaouzidis, D. K. Iakovidis, D. E. Yung, E. Rondonotti, U. Kopylov, J. N. Plevris, E. Toth, A. Eliakim, G. W. Johansson, and W. Marlicz, "KID Project: an internet-based digital video atlas of capsule endoscopy for research purposes," *Endoscopy international open,* vol. 5, pp. E477-E483, 2017.

[12] P. Coelho, A. Pereira, A. Leite, M. Salgado, and A. Cunha, "A Deep Learning Approach for Red Lesions Detection in Video Capsule Endoscopies," *Image Analysis and Recognition: In Proceeding of the 15th International Conference (ICIAR),* vol. 5, pp. 553-561, 2018.

[13] Q. Su, F. Wang, D. Chen, G. Chen, C. Li, and L. Wei, "Deep convolutional neural networks with ensemble learning and transfer learning for automated detection of gastrointestinal diseases," *Computers in Biology and Medicine,* vol. 150, p. 106054, 2022.

[14] K. Pogorelov, K. R. Randel, C. Griwodz, S. L. Eskeland, T. de Lange, D. Johansen, C. Spampinato, D.-T. Dang-Nguyen, M. Lux, and P. T. Schmidt, "Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection," in *Proceedings of the 8th ACM on Multimedia Systems Conference*, pp. 164-169, 2017.