

Supplementary Material

Artificial Intelligence-Based Analytics for Diagnosis of Small Bowel Enteropathies and Black Box Feature Detection

Appendix S1. Supplemental Methods – ResNet50

Appendix S2. Supplemental Methods – Custom Shallow CNN

Appendix S3. Supplemental Methods – Gradient-weighted Class Activation Mappings

Appendix S4. Supplemental Methods and Results – CellProfiler

Table S1: Modified Marsh score classification for Celiac Disease

Table S2: Error Analysis and performance statistics for deep learning models: modified ResNet50, ResNet50 with multi-zoom architecture, shallow CNN, and ensemble model.

Table S3: Regions of interest highlighted by Gradient-weighted Class Activation Mappings

(Grad-CAMs) that were obtained for all patches trained via modified ResNet50, ResNet50 Multi-zoom, and Shallow CNN. Key – EE: Environmental Enteropathy; CD: Celiac Disease; Controls: Histologically Normal Controls; IELs: Intra-epithelial Lymphocytes; LP: Lamina Propria

Figure S1. Biopsy Image Patch Creation – Images were split into patches of 1000x1000 and 2000x2000 pixels with an overlap (horizontal and vertical axis) of 750 and 1000 pixels, respectively.

Figure S2. Sample Augmentation – Biopsy image patches were randomly rotated (90, 180, 270 or 360 degree angles), mirrored, zoomed (between 1x and 1.1x).

Figure S3. Environmental Enteropathic Dysfunction Biopsy Initiative (EEDBI) scoring system for Environmental Enteropathy

Figure S4. CellProfiler – isolation of nuclei from high resolution biopsies via separation of colors and further thresholds do generate grayscale image for mapping of all nuclei present in a given image.

Supplementary References

Appendix S1. Supplemental Methods – ResNet50

ResNet50 is a widely used deep CNN architecture with 50 layers for image classification requiring identification of microscopic patterns. We modified the final decision layers of ResNet50 to improve accuracy¹. To combat data sparsity we used transfer learning, an established methods used to improve training using limited datasets by pre-training the model on the ImageNet dataset.(1, 2) Since different layers capture different information, discriminative fine tuning approach was utilized as described by Howard and Ruder.(3) We then trained a patch level modified ResNet50 architecture for classification of EE vs CD vs controls (Figure 2). Layers closer to the input were more likely to have learned more general features, while later layers identified more abstract features. The learning rate used for training the initial and middle layers was 1/9th and 1/3rd the rate of the final layers, respectively. Cyclic cosine annealing with restarts to prevent the model from getting stuck in local minima while training was use.(4) By lowering the learning rate periodically, we also nudged the model towards the global minima. Learning rate was intermittently reset (starting with a larger value) so it could move out of the local minima in case it got stuck and eventually reach the global optimum. By restarting, the need to experimentally estimate learning rate values was eliminated. Test Time Augmentation (TTA) was performed for final predictions to ensure they were not sensitive to image orientation. TTA randomly performs augmentations (zoom, tilt, brightness) on the images during prediction which enables the model to identify common micro level patterns with little regard for image orientation. The model was trained over 10 epochs with a batch size of 32.

Appendix S2. Supplemental Methods – Custom Shallow CNN

For the custom Shallow CNN, the input layer began with biopsy image patches which were sequentially fed into three convolution layers, each followed by a maxpool layer to reduce the dimensions of the input. The output from the last convolutional layer was then flattened and passed to a fully connected perception layer. The purpose of MaxPooling was to select the maximum element in the pooling window in order to reduce outputs while preserving important features. The output layer then contained three nodes for classification of images as either EE, CD or controls.

Appendix S3. Supplemental Methods – Gradient-weighted Class Activation Mappings

Gradient-weighted Class Activation Mappings (Grad-CAMs) generate the localization heatmaps highlighting specific regions of interes.(5) Activation values from an intermediate convolution layer and corresponding gradients as described by Selvaraju et al.(5) to generate these heatmaps.

Appendix S4. Supplemental Methods and Results – CellProfiler

An alternative method, using CellProfiler, for explainability of the classification models for models based on the duodenal biopsy images was explored. This method has previously been used for lung cancer biopsy images and it extracts cellular features from the images and further feeds them into machine learning models like random forests or support vector machines.(6) To achieve this, we used open source software CellProfiler which provides various modules to isolate nucleated cells in high resolution biopsies. We used UnmixColors module to separate colors and further used Otsu thresholds on the grayscale image generated to create a map of all nuclei present in a given image using IdentifyPrimaryObjects. Features were extracted on the patch level of size 1000x1000 and then averaged over the whole biopsy image. We used these features to feed into a random forest classifier to predict disease class, Environmental Enteropathy (EE) versus Celiac Disease (CD) versus histologically normal controls, and we achieved an accuracy of 65%. The dataset was split randomly into training and test sets to validate our accuracy. However, we observed that the model was correctly classifying CD but performed worse when classifying EE and controls. This showed that an approach such as this one may not be feasible for diseases similar to our dataset with histological overlap and a range of cellular features utilized for diagnosis and assessment.

Table S1: Modified Marsh score classification for Celiac Disease.(7)

Marsh Score	Intraepithelial Lymphocytes/100 enterocytes in duodenum	Crypt Hyperplasia	Villi
0	<30	Normal	Normal
1	>30	Normal	Normal
2	>30	Increased	Normal
3a	>30	Increased	Mild Atrophy
3b	>30	Increased	Marked Atrophy
3c	>30	Increased	Complete Atrophy

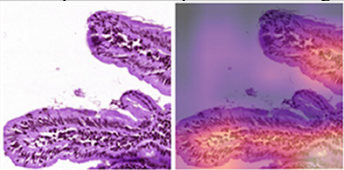
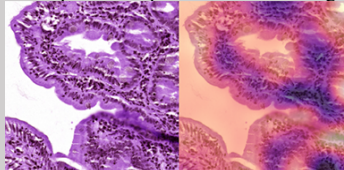
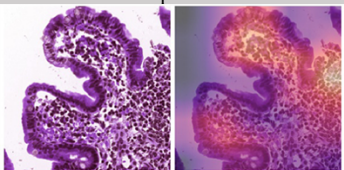
Table S2: Error Analysis and performance statistics for deep learning models: modified ResNet50, ResNet50 with multi-zoom architecture, shallow CNN, and ensemble model.

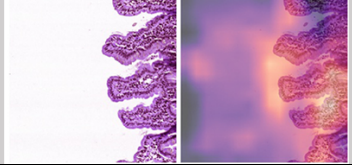
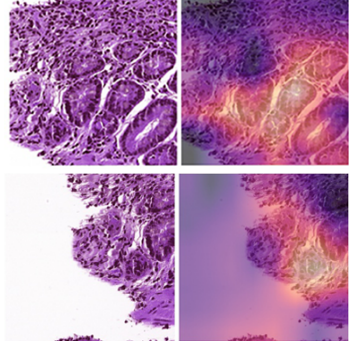
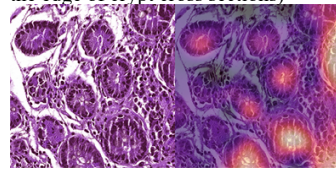
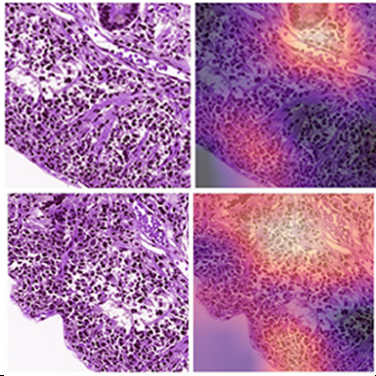
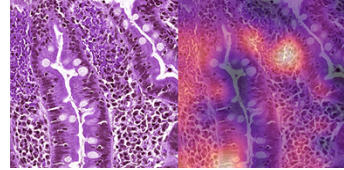
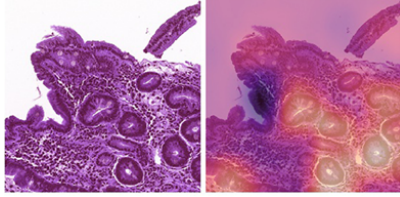
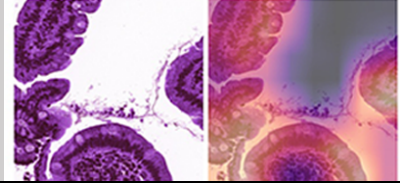
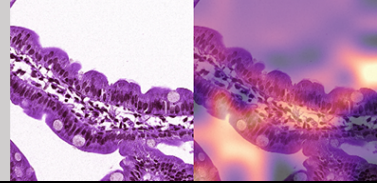
Key – EE: Environmental Enteropathy; CNN: Convolutional Neural Network; PPV: Positive Predictive Value; NPV: Negative Predictive Value

Deep Learning Model	Class	Sensitivity	Specificity	PPV	NPV	Precision	Recall	F1-score	Support
ResNet50	Celiac	0.80	1.00	1.00	0.82	0.98	0.86	0.92	51
	EE	1.00	1.00	1.00	1.00	1.00	1.00	1.00	11
	Normal	1.00	0.83	0.78	1.00	0.83	0.97	0.90	36
	<i>Average/ Total</i>	0.93	0.94	0.93	0.94	0.94	0.94	0.94	98
ResNet50 with multi-zoom	Celiac	0.92	0.98	0.98	0.92	0.98	0.86	0.92	51
	EE	1.00	1.00	1.00	1.00	1.00	1.00	1.00	9
	Normal	0.97	0.93	0.90	0.98	0.83	0.97	0.90	36
	<i>Average/ Total</i>	0.96	0.97	0.96	0.97	0.94	0.94	0.94	96
Shallow CNN	Celiac	0.92	0.71	0.78	0.89	0.78	0.92	0.85	51
	EE	0.78	1.00	1.00	0.98	1.00	0.78	0.88	9
	Normal	0.69	0.93	0.86	0.84	0.86	0.69	0.77	36
	<i>Average/ Total</i>	0.80	0.88	0.88	0.90	0.88	0.80	0.83	96
Ensemble	Celiac	0.90	0.96	0.96	0.90	0.96	0.90	0.93	51
	EE	1.00	1.00	1.00	1.00	1.00	1.00	1.00	9
	Normal	0.94	0.92	0.87	0.96	0.87	0.94	0.91	36
	<i>Average/ Total</i>	0.95	0.96	0.94	0.95	0.94	0.95	0.95	96

Table S3: Regions of interest highlighted by Gradient-weighted Class Activation Mappings (Grad-CAMs) that were obtained for all patches trained via modified ResNet50, ResNet50 Multi-zoom, and Shallow CNN.

Key – EE: Environmental Enteropathy; CD: Celiac Disease; Controls: Histologically Normal Controls; IELs: Intra-epithelial Lymphocytes; LP: Lamina Propria; CNN: Convolutional Neural Network

	ResNet50 (with and without Multi-zoom)	Shallow CNN
EE	<ul style="list-style-type: none"> Superficial Epithelium: high numbers of IELs 	<ul style="list-style-type: none"> Superficial Epithelium: IELs and goblet cells 
	<ul style="list-style-type: none"> LP: mononuclear cells, areas of white slit-like spaces which were artefactual separation of tissue 	<ul style="list-style-type: none"> The heatmaps focused on too many features in each patch due to which further features could not be categorized.
	<ul style="list-style-type: none"> Ignored crypt cross sections 	
	<ul style="list-style-type: none"> Multi-zoom: focused on the villus tips and similar LP areas as ResNet50 without Multi-zoom 	

		
CD	<ul style="list-style-type: none"> • Crypt Cross-sections: crypt epithelium with crowded nuclei (top image), edge of the crypts, and crushed and telescoping crypts (bottom image). 	<ul style="list-style-type: none"> • Inner lumen of the crypts (compared to ResNet50 focusing on the edge of crypt cross sections) 
	<ul style="list-style-type: none"> • LP: intervening LP between adjacent crypts, lymphocytes stacked in a row simulating a linear epithelial architecture (top image), and areas with cells consistent with mononuclear cellular infiltrate (bottom image). 	<ul style="list-style-type: none"> • LP: mononuclear cell areas 
	<ul style="list-style-type: none"> • Similar to EE: the model also focused on mononuclear cell infiltrate in the LP but mainly included areas adjacent to crypts. In some patches surface epithelium with high IELs were highlighted 	
	<ul style="list-style-type: none"> • Multi-zoom: similar areas as ResNet50 were mapped with a primary focus on crypts and the areas adjacent to them. ResNet50 Multi-zoom focused on several crypts rather than 1 to 2 crypts compared to ResNet50. 	
Controls	<ul style="list-style-type: none"> • Superficial Epithelium: tall columnar cells and brush border (highly depictive of normal epithelium(8)). 	<ul style="list-style-type: none"> • Surface Epithelium: epithelial cells containing abundant cytoplasm and goblet cells. 

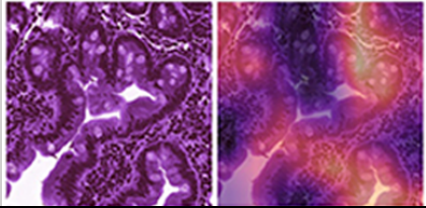
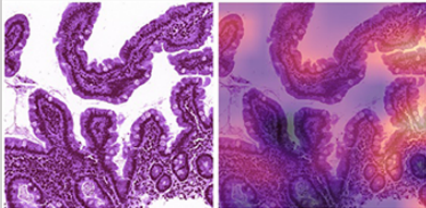
	<ul style="list-style-type: none"> • Crypt bases resembling superficial epithelium 	
	<ul style="list-style-type: none"> • Similar to EE: areas with artefactual separation of tissue in the LP or adjacent to the crypt cross sections 	
	<ul style="list-style-type: none"> • Multi-zoom: similar to ResNet50 including areas containing superficial epithelium with abundant cytoplasm (columnar/ brush border features not as highly appreciated at this zoom level) and rich goblet cells. 	

Figure S1. Biopsy Image Patch Creation – Images were split into patches of 1000x1000 and 2000x2000 pixels with an overlap (horizontal and vertical axis) of 750 and 1000 pixels, respectively.

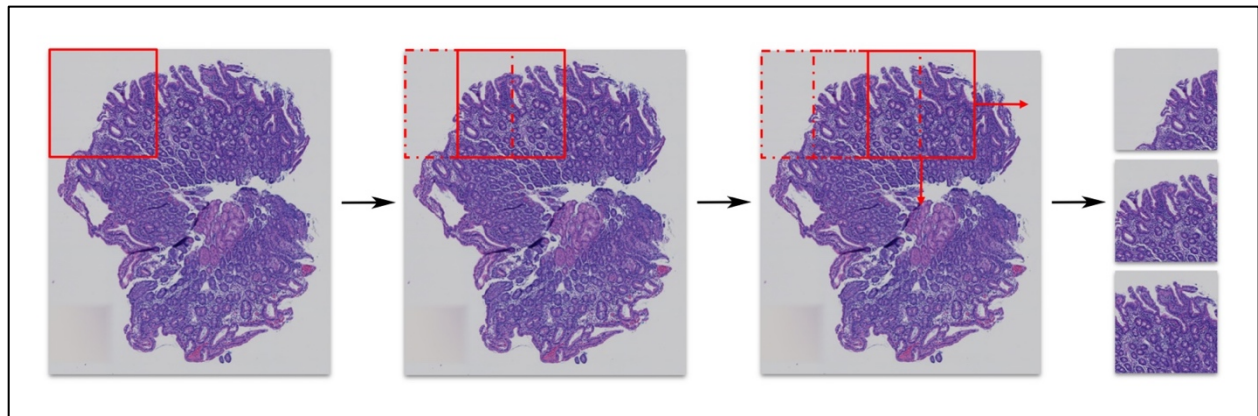


Figure S2. Sample Augmentation – Biopsy image patches were randomly rotated (90, 180 or 270 degree angles), mirrored, zoomed (between 1x and 1.1x).

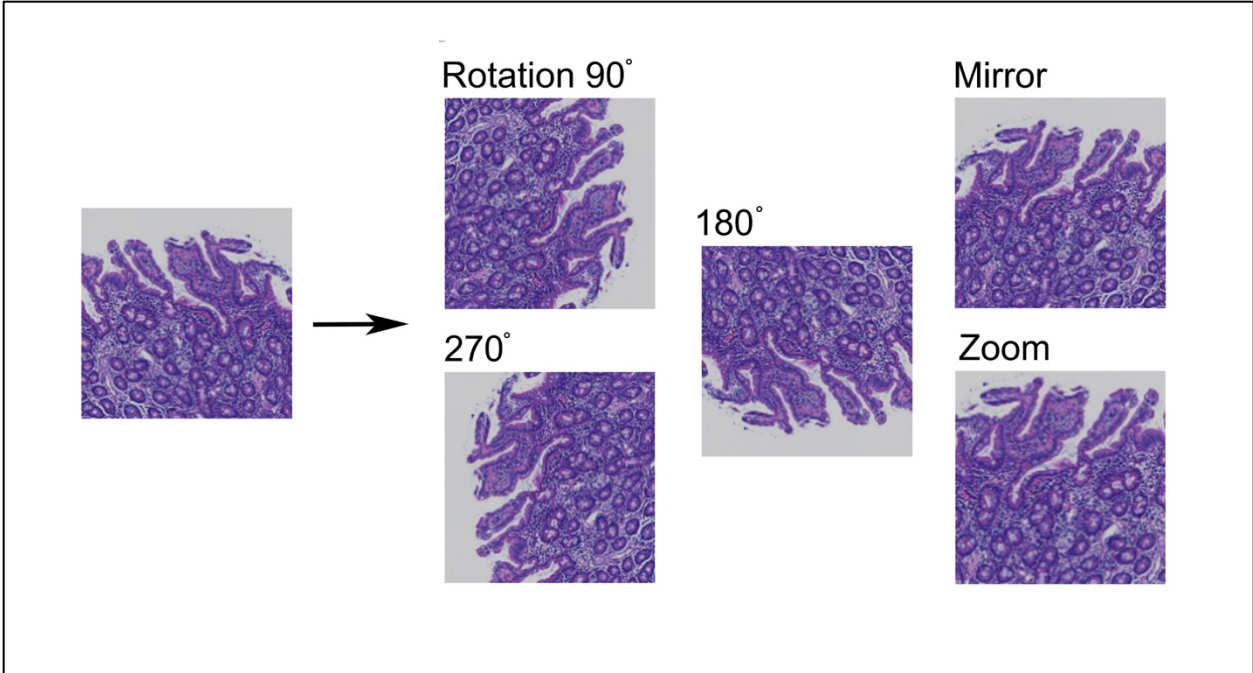
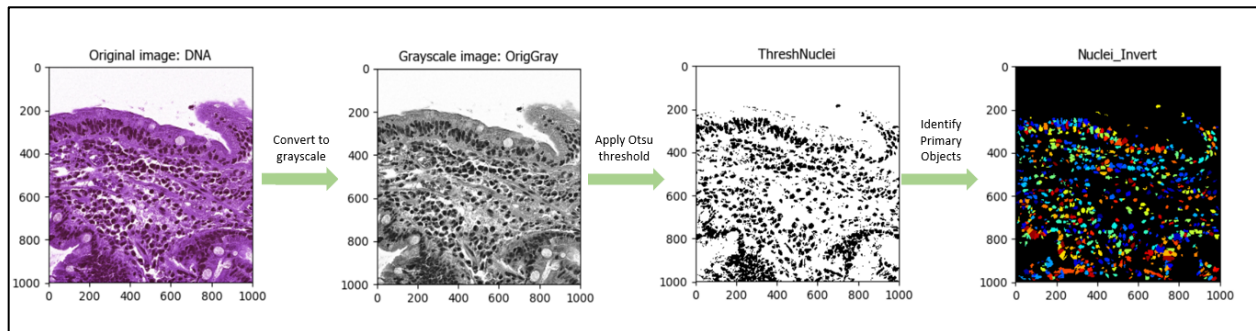


Figure S3. Environmental Enteropathic Dysfunction Biopsy Initiative (EEDBI) scoring system for Environmental Enteropathy.(9)

Acute (neutrophilic) inflammation	0: PMNs may be present in vessels or in lamina propria, but there is no intraepithelial infiltration (cryptitis, villitis)
	1: 1–2 foci of epithelial neutrophilic infiltration or crypt microabscesses
	2: > 2 foci of epithelial neutrophilic infiltration or crypt microabscesses, but ≤ 50% of mucosa involved
	3: > 50% of mucosa involved by epithelial neutrophilic infiltration
Eosinophil infiltration	0: No increase in eosinophils (highly scattered in lamina propria, no intravillous or intercryptal space with > 5 eosinophils)
	1: Increased eosinophils (intravillous or intercryptal space with > 5 eosinophils) involving ≤ 50% of mucosa, with no eosinophilic crypt microabscesses
	2: Increased eosinophils (intravillous or intercryptal space with > 5 eosinophils) involving > 50% of mucosa, or up to 1 focus of eosinophilic epithelial infiltration or crypt microabscesses per mucosal fragment
	3: > 2 foci of eosinophilic epithelial infiltration or crypt microabscesses in any mucosal fragment
Chronic inflammation	0: No qualitative increase in mononuclear inflammatory cells (MIC) in lamina propria. Majority of villous bases contain < 3 MIC across, on average.
	1: Increased MIC, based on villous base displaying 3–5 MIC across, on average.
	2: Increased MIC, based on villous base displaying 6–10 MIC across, on average.
	3: Increased MIC, based on villous base displaying > 10 lymphocytes on average.
Intraepithelial lymphocytes	0: IEL ratio of lymphocytes to enterocytes < = 20% in any area
	1: Lymphocyte/epithelial ratio > 20%, but ≤ 50%, in < = 50% of mucosa
	2: Lymphocyte/epithelial ratio > 20%, but ≤ 50%, in > 50% of mucosa
	3: Lymphocyte/epithelial ratio > 50% in ≤ 50% of mucosa
	4: Lymphocyte/epithelial ratio > 50% in > 50% of mucosa
Villous architecture	0: Majority of villi are > 3 crypt lengths long
	1: Villi are ≤ 3 but > 1 crypt length long, with abnormality in ≤ 50% of mucosa.
	2: Villi are ≤ 3 but > 1 crypt length long, with abnormality in > 50% of mucosa
	3: Villi absent, or ≤ 1 crypt length long, with abnormality in ≤ 50% of mucosa
	4: Villi absent, or ≤ 1 crypt length long, with abnormality in > 50% of mucosa
Intramucosal Brunner glands	0: Brunner glands are in submucosa, but are not observed above the muscularis mucosae
	1: 1–2 foci, none involving > 5 crypt bases
	2: 3–5 foci, none involving > 5 crypt bases
	3: > 5 foci, or any area of intramucosal Brunner glands involving > 5 crypt bases
Foveolar cell metaplasia	0: Only absorptive enterocytes and goblet cells observed on villi, no evidence of foveolar cells
	1: Foveolar mucin cells observed, usually on the tips of the villi; 1–2 villous tips involved
	2: Foveolar mucin cells observed, usually on the tips of the villi; 3–5 villous tips involved
	3: Foveolar mucin cells observed, usually on the tips of the villi; > 5 villous tips involved
Goblet cell density	0: Normal goblet cell density (at least 1 goblet cell per 20 enterocytes) in all evaluable mucosal epithelial layer
	1: Decreased goblet cells (< 1/20 enterocytes) in 1–25% of evaluable mucosal epithelium
	2: Decreased goblet cells (< 1/20 enterocytes) in 26–50% of evaluable mucosal epithelium
	3: Decreased goblet cells (< 1/20 enterocytes) in 51–75% of evaluable mucosal epithelium
	4: Decreased goblet cells (< 1/20 enterocytes) in 76–100% of evaluable mucosal epithelium
Paneth cell density	0: ≥ 5 Paneth cells/ crypt, on average
	1: 2–4 Paneth cells/ crypt, on average
	2: < 2 Paneth cell/crypt, involving ≤ 50% of crypt bases
	3: < 2 Paneth cell/crypt, involving > 50% of crypt bases
Enterocyte injury	0: Majority of enterocytes (90%) show tall columnar morphology
	1: Enterocytes show low columnar (< 2:1 L:W ratio), cuboidal or flat morphology, in ≤ 50% of mucosa
	2: Enterocytes show low columnar (≤ 2:1 L:W ratio), cuboidal or flat morphology, in > 50% of mucosa
	3: Any area of mucosal erosion/ulceration
Epithelial detachment	0: Complete coverage of mucosal surface by epithelial cells
	1: Surface epithelium missing or detached from < 25% of mucosa
	2: Surface epithelium missing or detached from 25–50% of mucosa
	3: Surface epithelium missing or detached from 51–75% of mucosa
	4: Surface epithelium missing or detached from > 75% of mucosa
Note: For chronic inflammation there is a response option “not scorable due to <3 villi assessable or other slide quality issue”, while for all other parameters there is a response option “not scorable due to slide quality or other factors”.	
https://doi.org/10.1371/journal.pntd.0007975.t003	

Figure S4. CellProfiler – isolation of nuclei from high resolution biopsies via separation of colors and further thresholds do generate grayscale image for mapping of all nuclei present in a given image. Types of image features usually included are cell size, shape, distribution of pixel intensity in the cells and nuclei, as well as texture of the cells and nuclei



Supplementary References

1. Pan SJ, Yang Q. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*. 2009;22(10):1345-59.
2. Deng J, Dong W, Socher R, et al., editors. Imagenet: A large-scale hierarchical image database. 2009 IEEE conference on computer vision and pattern recognition; 2009: Ieee.
3. Howard J, Ruder S. Universal language model fine-tuning for text classification. *arXiv preprint arXiv:180106146*. 2018.
4. Loshchilov I, Hutter F. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:160803983*. 2016.
5. Selvaraju RR, Cogswell M, Das A, et al., editors. Grad-cam: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE International Conference on Computer Vision*; 2017.
6. Yu K, Zhang C, Berry G, et al. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat Commun*. 2016; 7: 12474. Epub 2016/08/17. <https://doi.org/10.1038/ncomms12474> PMID: 27527408.
7. Modified Marsh Classification of histologic findings in celiac disease (Oberhuber) Stanford Medicine [Available from: <http://surgpathcriteria.stanford.edu/gi/ceciac-disease/marsh.html>].
8. Serra S, Jani PA. An approach to duodenal biopsies. *Journal of clinical pathology*. 2006;59(11):1133-50.
9. Liu T-C, VanBuskirk K, Ali SA, et al. A novel histological index for evaluation of environmental enteric dysfunction identifies geographic-specific features of enteropathy among children with suboptimal growth. *PLoS neglected tropical diseases*. 2020;14(1):e0007975.