

Detection of Intestinal Protozoa in Trichrome-Stained Stool Specimens by Use of a Deep Convolutional Neural Network

Blaine A. Mathison,^a Jessica L. Kohan,^a John F. Walker,^b Richard Boyd Smith,^b Orly Ardon,^{a,c} Marc Roger Couturier^{a,c}

^aARUP Laboratories, Salt Lake City, Utah, USA

^bTechcyte, Inc., Lindon, Utah, USA

^cDepartment of Pathology, University of Utah School of Medicine, Salt Lake City, Utah, USA

Orly Ardon and Marc Roger Couturier contributed equally as investigators. Author order was determined by increasing seniority.

ABSTRACT Intestinal protozoa are responsible for relatively few infections in the developed world, but the testing volume is disproportionately high. Manual light microscopy of stool remains the gold standard but can be insensitive, time-consuming, and difficult to maintain competency. Artificial intelligence and digital slide scanning show promise for revolutionizing the clinical parasitology laboratory by augmenting the detection of parasites and slide interpretation using a convolutional neural network (CNN) model. The goal of this study was to develop a sensitive model that could screen out negative trichrome slides, while flagging potential parasites for manual confirmation. Conventional protozoa were trained as “classes” in a deep CNN. Between 1,394 and 23,566 exemplars per class were used for training, based on specimen availability, from a minimum of 10 unique slides per class. Scanning was performed using a 40× dry lens objective automated slide scanner. Data labeling was performed using a proprietary Web interface. Clinical validation of the model was performed using 10 unique positive slides per class and 125 negative slides. Accuracy was calculated as slide-level agreement (e.g., parasite present or absent) with microscopy. Positive agreement was 98.88% (95% confidence interval [CI], 93.76% to 99.98%), and negative agreement was 98.11% (95% CI, 93.35% to 99.77%). The model showed excellent reproducibility using slides containing multiple classes, a single class, or no parasites. The limit of detection of the model and scanner using serially diluted stool was 5-fold more sensitive than manual examinations by multiple parasitologists using 4 unique slide sets. Digital slide scanning and a CNN model are robust tools for augmenting the conventional detection of intestinal protozoa.

KEYWORDS protozoa, ova and parasite exam, artificial intelligence, machine learning, digital microscopy, convolutional neural network, parasites, trichrome stain

Intestinal protozoa are responsible for relatively few infections in the developed world but for which the testing volume is disproportionately high. Manual microscopic evaluation of stool, the ova-and-parasite (O&P) examination, is still considered the “gold standard” detection method after nearly a century of use (1). This method suffers from variable sensitivity (protozoa, operator, and laboratory dependent), is time- and resource-consuming, and represents one of the true “lost or dying arts” of traditional microbiology. Maintaining staff competency and engagement are significant challenges for the clinical parasitology laboratory. The clinical parasitology laboratory further suffers from two major aspects of workforce-related challenges: first, recently educated technologists increasingly gravitate toward technology-driven, automated disciplines of laboratory medicine (e.g., mass spectrometry, next-generation nucleic acid sequencing, massively multiplexed pathogen detection, and specimen-to-answer automated testing); and second, there is a lack of adequately trained or available

Citation Mathison BA, Kohan JL, Walker JF, Smith RB, Ardon O, Couturier MR. 2020. Detection of intestinal protozoa in trichrome-stained stool specimens by use of a deep convolutional neural network. *J Clin Microbiol* 58:e02053-19. <https://doi.org/10.1128/JCM.02053-19>.

Editor Bobbi S. Pritt, Mayo Clinic

Copyright © 2020 American Society for Microbiology. All Rights Reserved.

Address correspondence to Marc Roger Couturier, marc.couturier@aruplab.com.

For a commentary on this article, see <https://doi.org/10.1128/JCM.00511-20>.

Received 13 December 2019

Returned for modification 10 February 2020

Accepted 6 April 2020

Accepted manuscript posted online 15 April 2020

Published 26 May 2020

personnel (2). Despite nonmicroscopic advancements in the form of antigen or molecular detection of intestinal protozoa, as well as efforts to streamline fixative/collecion and processing of O&P specimens (3), technological advancements and efforts to augment O&P detection and interpretation have been seemingly absent in this field.

The traditional O&P examination in many countries consists of a concentrated wet mount to detect helminth eggs/larvae and protozoan cysts and a permanently stained trichrome slide for the detection of protozoan cysts and trophozoites. Advancements in health care, infrastructure, and sanitation over the last century has resulted in fewer intestinal protozoal infections in the United States. Most cases seen are from immigrants or travelers to areas of endemicity. As such, parasite morphologists can spend a large percentage of their time screening negative specimens (e.g., ~95% to 98% of specimens in our large reference parasitology laboratory), which can result in repetitive stress injuries, low job satisfaction, and diagnostic errors due to fatigue or inexperience (4). This discipline is ripe for application of an augmentation process for the microscopic examination component of the method.

Detection and specification of *Plasmodium* sp. in blood using digital microscopy has been investigated by multiple groups to date (5–7); however, there is a dearth of research dedicated toward application of digital microscopy for intestinal parasites. Digital microscopic detection of protozoa from a complicated/heterogeneous matrix, such as stool, represents a significant technical/scientific barrier to overcome compared with more homogenous and fluidic matrices. Only a few preliminary, proof-of-concept studies have been reported, which aim to improve the detection of helminth eggs (8–13) and protozoan cysts (11) in human stool specimens, but none of these have been applied for routine clinical diagnosis. To date, there have been no significant technological advancements for the detection of protozoa in human stool specimens using permanently stained slides (e.g., trichrome, modified acid-fast, and modified safranin).

It is difficult for traditional computer vision algorithms to detect parasites on trichrome-stained fecal samples because parasites are embedded in debris from numerous organic shapes from plants, food contents, and other microbiota. The human process of scanning quickly for organic shapes and then making a separate careful observation of morphological features (size, shape, and internal and external features) is simultaneous and best modeled using a deep-learning-based, convolutional neural network-based (CNN) model. This study aimed to develop a CNN model, paired with high-resolution digital slide scanning, to detect common intestinal protozoa in human stool specimens stained with trichrome. The work was segmented into three phases: (i) collect and digitally scan well-defined trichrome-stained specimens from our reference parasitology laboratory containing various targeted species and morphological stages of protozoa (classes), (ii) feed the aggregate digital image data into the CNN-model to train it to recognize defined classes, and (iii) perform clinical laboratory validation of the resulting model for use in a licensed diagnostic parasitology laboratory. This resulting model and clinical laboratory validation serves as an augmentation to the current manual microscopic method, allowing a streamlined review process for all digitally evaluated slides.

MATERIALS AND METHODS

Classification of categories for model development. Training classes were identified in an effort to comprehensively detect necessary targets reported by standard trichrome staining; those classes included *Giardia duodenalis* cysts and trophozoites, *Entamoeba hartmanni* trophozoites, *Entamoeba* sp. non-*hartmanni* (i.e., the “large” *Entamoeba* sp.) trophozoites, *Dientamoeba fragilis*, *Blastocystis* species, *Chilomastix mesnili* trophozoites, *Endolimax nana*/*Iodamoeba buetschlii* trophozoites, red blood cells (RBCs), and white blood cells (WBCs). The software was also trained to recognize yeast as an anticlass to prevent class confusion with other categories. For *Entamoeba* spp., *C. mesnili*, *E. nana*, and *I. buetschlii*, the software was only trained on labels that represented the morphologically distinct trophozoites. Cysts for those organisms were not trained in the model due to a low number of high-quality exemplars and poor quality of morphology on the trichrome stain. For the non-*hartmanni* *Entamoeba* species, the training was performed using the characteristic nucleus and chromatin dot as the labeled feature.

Specimen collection, preparation, and scanning. One-hundred twenty-seven slides that were previously reported as positive in the diagnostic laboratory were used for training the software. Each

TABLE 1 Total number of unique slides per class, and total number of examples per class used for training the model

Category (class)	No. of unique slides per class	No. of examples per class
<i>Giardia duodenalis</i> cyst	23	6,499
<i>Giardia duodenalis</i> trophozoite	21	2,191
<i>Blastocystis</i> sp.	61	23,566
<i>Dientamoeba fragilis</i>	29	12,764
<i>Entamoeba non-hartmanni</i> trophozoite	34	4,307
<i>Entamoeba hartmanni</i> trophozoite	10	1,394
<i>Chilomastix mesnili</i> trophozoite	15	4,064
<i>Endolimax nana</i> / <i>Iodamoeba buetschlii</i> trophozoite	36	7,914
Red blood cells	18	8,482
White blood cells	31	2,099
Yeast	94	13,450

category and their numbers of unique specimens (slides) are as follows: *G. duodenalis*, cyst ($n = 23$); *G. duodenalis*, trophozoite ($n = 21$); *Blastocystis* sp. ($n = 61$); *D. fragilis* ($n = 29$); *E. hartmanni* ($n = 10$); *Entamoeba* spp., non-*hartmanni* ($n = 34$; species trained included *Escherichia coli*, $n = 21$; *Entamoeba histolytica*/*Entamoeba dispar*, $n = 10$; *Entamoeba polecki*; $n = 1$; *Entamoeba* sp., [not otherwise specified]; $n = 3$); *C. mesnili* ($n = 15$); *E. nana*/*I. buetschlii* ($n = 36$); RBCs ($n = 18$); WBCs ($n = 31$); and yeast ($n = 94$). At the time of clinical laboratory validation (see below), each of the classes in training and their numbers of labels were as follows: *G. duodenalis*, cyst ($n = 6,499$); *G. duodenalis*, trophozoite ($n = 2,191$); *Blastocystis* sp. ($n = 23,566$); *D. fragilis* ($n = 12,764$); *Entamoeba hartmanni* trophozoite ($n = 1,394$); *Entamoeba* sp. non-*hartmanni* trophozoite ($n = 4,307$); *C. mesnili* trophozoite ($n = 4,064$); *E. nana*/*I. buetschlii*, trophozoite ($n = 7,914$); RBCs ($n = 8,482$); WBCs ($n = 2,099$); and yeast ($n = 13,450$) (Table 1). Slides were chosen from specimens preserved in a variety of fixatives, including polyvinyl alcohol, sodium acetate-acetic acid-form (SAF), and several single-vial alcohol-based preservatives.

All slides were coverslipped prior to scanning, either manually using Permount (Fisher Scientific, Hampton, NH) or by using an automated coverslipper (Tissue-Tek film; Sakura Finetek, Torrance, CA). Slides were imaged using a Panoramic 250 Flash III (3DHISTECH, Budapest, Hungary) equipped with a $\times 40$ magnification objective (0.95 numerical aperture) and an optical doubler, resulting in an $\times 82.4$ magnification image with a resolution of $0.1214 \mu\text{m}$ per pixel. Fields were scanned at three different layers and the scanner software selected the best focal plane from the Z-stack, and the scanned fields were stitched together to form the complete scanned image. An acceptable scan was defined as having approximately 80% of the slide in focus and analyzed by software. An unacceptable scan was blurry or if greater than approximately 20% of the slide did not scan. A failed scan did not scan at all.

Initial classification and labeling. Scanned images were manually reviewed for candidate organisms that clearly displayed defining features of organisms using Panoramic Viewer software (3DHISTECH). The candidate organisms were “labeled” by creating a tight box around the whole organism and specifying the organism type. The manually labeled organisms constituted the initial training sets for each class of organism. Manual candidate labeling occurred on known positive slides for each class. Manual labeling is very labor-intensive and unlikely to find every organism on a given slide. To eliminate the need for each slide to be manually searched, the trained model was used to classify the manually labeled slides to search for additional candidate organism labels (known as “find boxes”). Find boxes allowed the software to search for organisms with similar features to the classes in the initial training sets, which were then verified by a human expert before being used in future training runs. See Fig. S1 in the supplemental material for flow chart details of this process.

Initial and new labels were evaluated and supervised by an expert parasitologist (with over 20 years of clinical parasitology expertise, including 9 years of training and employment at the Centers for Disease Control and Prevention diagnostic parasitology laboratory) for accuracy and quality, and inaccurate labels were corrected. Labels on an indiscernible organism (e.g., poor quality, bad focus, or no discernible features) were labeled “excluded” and not included in the training set. Proposed labels that represented artifacts (or other objects not trained on) were relabeled as “background” and used as “negative examples” in future trainings. Incorrect candidate labels on a discernible organism were manually reclassified to the correct organism. Proposed labels also had their boxes corrected where required to wholly contain but tightly encompass the organism. The “find new boxes” process was repeated to increase the number of exemplars per class until find boxes began failing to find new exemplar labels or a count of 300 to 400 exemplars from any individual slide was reached.

Training a deep CNN. During development, numerous training runs were executed and the resulting metrics (see “Analysis of model performance on a per-label basis”) used to evaluate progress (full training iteration flowchart in Fig. 1). The training data set is made of all labels from all classes, including a number of representative background labels. “Scenes” from the training data set are generated dynamically every epoch for which augmentation is applied. A randomly arranged 250- by 250-pixel image was cropped to encompass a labeled box. Before beginning the training, 10% of the labels for each class were randomly selected and used as CNN training validation data to measure progress during and at the conclusion of training. The object detection model archi-

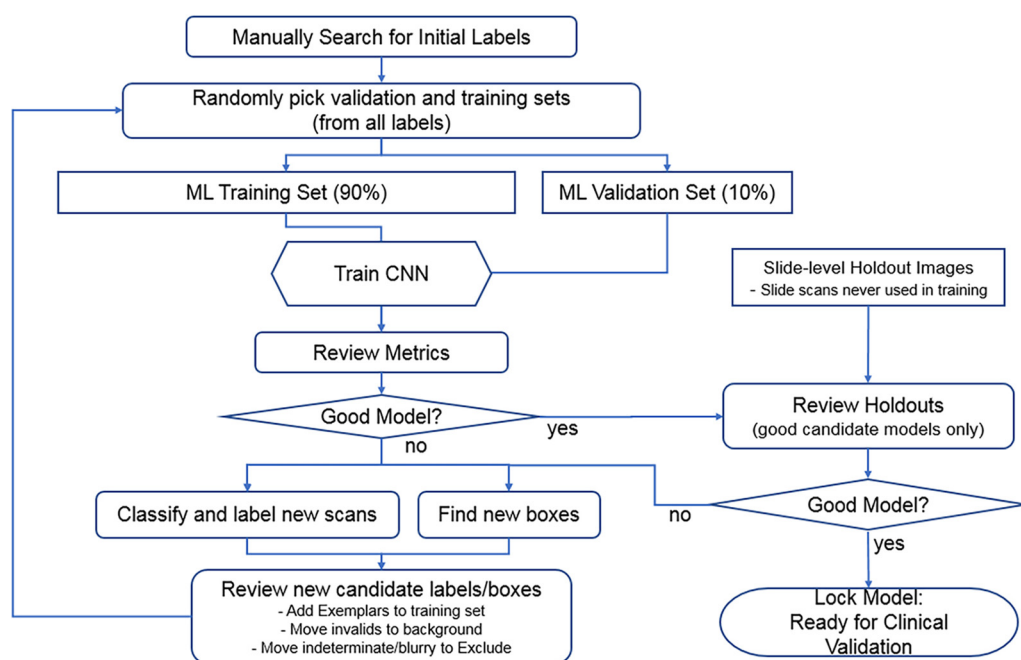


FIG 1 Flow chart summary of training iterations for CNN. ML, machine learning; CNN, convolutional neural network.

texture is a three color channel CNN based upon SSD Inception v2. The initial model was pretrained with the COCO image database (https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/detection_model_zoo.md; accessed November 6, 2017). All layers in the model were trained, including the convolutional layers using the object detection API produced by Google (https://github.com/tensorflow/models/tree/master/research/object_detection). This API leverages the TensorFlow library (v1.5.0; <https://www.tensorflow.org/>) and Keras (v2.2.4; <https://keras.io/>) for CNN training and execution.

Labeling was performed using the Techcyte (Techcyte, Inc., Linden, UT) cloud-hosted data storage and Web interface. For training, mini-batch gradient descent (batch size, 24) with Nesterov momentum (momentum, 0.9) (14) and cross entropy as the loss function (15) was used. A learning rate of 0.004 was used at initiation with decay at an exponential rate of 0.5 per epoch, starting from the second epoch. Each training run was 10 epochs.

Class balancing. The training data were constructed from a set of scanned sample images with labeled image boxes. This resulted in a variable number of available training images per class. This was an unavoidable consequence of limited access to specimens containing rare parasites. To force the CNN to learn all classes, the number of training images per class was normalized so that the images of the rarer classes were shown extra times as needed (with augmentation) to make each training epoch show the same number of each class for each epoch (e.g., if there were 3 classes to train with, A = 10,000, B = 15,000, and C = 8,000 labeled examples and 100,000 backgrounds for each epoch would contain randomly selected scenes for 15,000 As, 15,000 Bs, and 15,000 Cs by reusing 5,000 As and 7,000 Cs in different random 250- by 250-pixel cropped scenes from the host image).

During the creation of the training data, a large number of labeled image boxes for background were created. They were also used during each training epoch to provide “negative training” examples. Rather than normalizing all positive classes to the much larger number of background classes, class balancing was used during training with a 3:1 ratio of background to nonbackground exemplars/class examples (e.g., reusing the above example, normalizing 15,000 positive scenes per class would result in 45,000 background scenes being included in training).

Image processing for model training. Trichrome-stained, coverslipped slides were scanned (as above) and the images uploaded to the Techcyte cloud. The resulting image was approximately 13.1 mm by 10.2 mm, resulting in an 108,000- by 84,400-pixel image at $\times 82.4$ magnification with 0.1214 μm per pixel. The trained CNN was shown the sequence of 250- by 250-pixel scenes to look for parasites, for which it created a labeled image box. The organism may represent 3% to 25% of the pixel area of the scene for most organisms. Scenes were overlapped (horizontal and vertical) to prevent parasites being “sliced” at image boundaries and missed. Candidate labels with a confidence score below a cutoff threshold were rejected and not used. After the entire image was processed, duplicates from scene overlap were removed, and labeled image boxes were binned by class and sorted by decreasing confidence for display.

Augmentation. To reduce the number of unique labels required to train the model, random 250- by 250-pixel crops were used from the source image (all containing the label unsliced). This was done multiple times per label with different surrounding pixels and every such scene was randomly aug-

mented. This scene was taken from its source image to accurately display the example in its context environment and present the example as it would be seen during the nominal 250- by 250-pixel sample processing. The CNN was trained to process 250- by 250-pixel scenes. All of the labeled examples were smaller (in pixels) than this size.

Analysis of model performance on a per-label basis. Using the trained CNN, the labels were used to select random 250- by 250-pixel scenes to classify through the model, similar to how the training scenes were selected in the training data. The correct result was known so the scene could be evaluated for true positive/negative and detection of the box label with an intersection over union (IOU) ratio of 0.7. Correctly boxing and classifying the label was considered a true positive. A false-negative label was defined as no proposed box, proposing a box with the wrong class, or a label that could not be classified above the minimum confidence threshold (0.2). Machine-learning validation also included background labels for which the correct answer was equal to having no class above the minimum “confidence score” threshold. Precision-recall (P-R) curves were generated for both overall performance and on a per-class basis. The algorithm for P-R curves is to take all classified labels, sort by confidence score, and then track the evolution of machine learning precision (precision-ML) and recall as each classified label is added. Recall and precision-ML were modeled as both P-R curves and receiver operator characteristic (ROC) curves for each classification label. Recall was defined as true positive/(true positive + false negative). Precision-ML was defined as true negative/(true negative + false positive). Note that during classification, the ROC curve was not scored using every 250- by 250-pixel scene but only scenes known to contain items of interest (class-containing scenes or background). This means our ROC curves were not artificially inflated by a super majority of easy true negative scenes.

Clinical laboratory validation. Scan area design. The current laboratory practice of manual microscopy required a review of a minimum of 100 fields of view (FoV). The area needed to replicate the lab's FoV requirements was calculated and the measurements were applied in a scanning profile (3.8 mm by 10.2 mm). This optimized scanning area was intended to maximize detection of rare organisms, mimic clinical practices, and achieve a scan time between 4 and 5 minutes (specimen dependent). Progressive testing of scan areas resulted in altered shape, size, and Z-layer needed to achieve optimal target detection, scan quality, and time per scan. A titration series of positive slides was used to verify these performance metrics using the defined scan profile.

Slide (organism) classification algorithm. The algorithm was designed to prevent false negatives with minimal false-positive labels to increase the overall slide-level sensitivity. Proposed labels were presented to the operator in order of decreasing confidence, and true positive labels were confirmed. The resulting procedure was as follows: first, a trichrome-stained slide was scanned and automatically uploaded for processing; second, the uploaded image was processed as described in “Image processing for model training”; and third, the user views the labeled image boxes grouped by parasite class. In the third step, the user noted which parasite classes contained one or more true examples, and if no true examples were observed, the result was “negative.” If one or more true examples were observed, the result was “positive for class X.”

Specimen collection and preparation. Ten unique patient slides, previously reported as positive by standard O&P examination, were scanned and analyzed for each of the following eight categories: (i) *G. duodenalis*, (ii) *Blastocystis* species, (iii) *E. hartmanni*, (iv) *D. fragilis*, (v) *E. nana/l. buetschlii*, (vi) *C. mesnili*, (vii) red blood cells and white blood cells (mixed, in 3 to 4+ quantity for each), and (viii) mixed protozoa. Eleven unique patient slides for *Entamoeba* sp., non-*hartmanni* were scanned and analyzed. A cumulative total of 91 positive slides were used to establish accuracy during clinical laboratory validation. None of these slides were previously used for development or training. Single-organism categories may have also contained *Blastocystis* sp., as it is often found in the presence of other protozoa and finding unique single-organism infections can be very difficult. The “mixed protozoa” category was designated specifically for specimens that (i) contained two species, neither of which were *Blastocystis*, or (ii) contained three or more species, one of which could be *Blastocystis*. One-hundred twenty-five slides previously reported as negative by technologists in the clinical laboratory were also scanned and analyzed. All slides were coverslipped prior to scanning. Specimens were previously fixed in a variety of fixatives regularly received in our reference laboratory.

Slide-level accuracy. Slides were loaded into the scanner randomly and not organized by organism category to ensure that image analysis was performed in an unbiased manner. Analysis was performed using slide-level agreement (positive or negative for a parasite) as an accuracy metric. A true positive was defined as a slide containing parasites that the software detected. A true negative was defined as a slide that did not contain parasites and for which the model generated no labels or few labels that did not trigger a manual examination of the slide by a trained technologist. A false positive was defined as a slide that did not contain parasites but for which the model detected putative parasites that subsequently required manual review of the slides to refute. A false negative was defined as a slide that contained a parasite, but the parasite was not detected by the model.

Limit of detection. To test the limit of detection (LOD), a pooled stool specimen containing *Giardia duodenalis* and *Blastocystis* sp. was serially diluted in Alcorfix (Apacor, Wokingham, UK) mixed with human stool that was previously tested as negative for parasites by O&P. Four unique dilution series were prepared from the single pooled specimen. Two slides sets from each dilution series were prepared, namely one for scanning and one for manual microscopy. The series of 10 dilutions was divided between two O&P runs to disperse the positive specimens across multiple negative specimens in an attempt to avoid bias or suspicion.

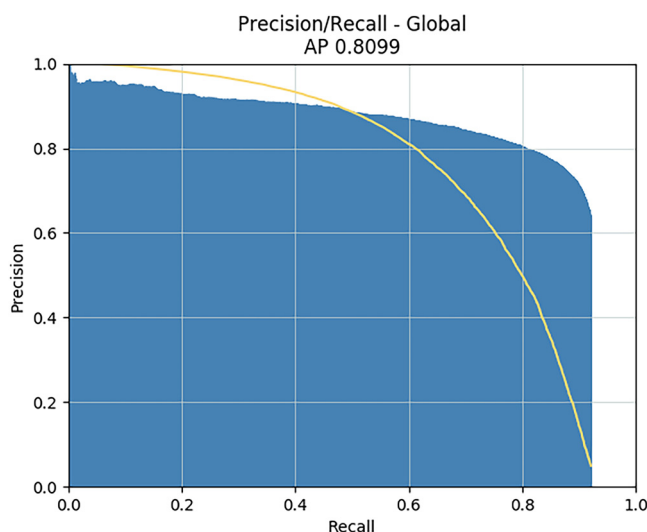


FIG 2 Global precision-recall curve for all classes combined from the CNN model. The yellow line shows the associated confidence score as each new label (both true positive and false positive) is added. The confidence score line is plotted on the same axis (the yellow line on the plots) as precision-ML to understand the evolution of recall with confidence. Image generated using Python Pillow library (<https://python-pillow.org/>).

RESULTS

Evaluation of model performance on a per-labeled image box basis.

Precision-recall plots. Due to the highly imbalanced nature of the data, we used the precision-recall plot to visualize model performance (16). The plot as shown in Fig. 2 is the P-R plot for all classes combined with the final model. This shows the incremental contribution of each new labeled image box from a confidence score-sorted list of labels found and classified by the newly trained model compared against the known ground truth labels of the machine learning validation data. “AP” is the average precision-ML up to the 0.05 confidence score cutoff. An example of constructing the P-R plot is provided in the supplemental material (see precision-recall plot construction).

The total recall achieved was approximately 92%, but this required accepting confidence scores down to 0.05, where the precision-ML was as low as ~65%. Applying a 0.4 confidence score cutoff (see below) yielded a recall of ~83% and precision-ML of ~78% (e.g., 3 of 4 labels selected are likely to be true). Over 60% of true positives (TPs) were detected before the confidence score fell below 0.8, while the FP did not increase above ~14%.

Confidence score cutoff selection and confidence class chart. As stated in “Scan/image processing,” the candidate-labeled image boxes found during CNN processing of the image included a confidence score. As the confidence score decreased, the CNN indicated a weaker match to the class selection of the candidate label, but the closest label would be given even if its score approached zero. Thus, as the confidence score decreased, the likelihood of the candidate label being a false positive increased. For this reason, the use of a confidence score cutoff was desirable to minimize the number of false-positive labels shown to the user without setting the confidence cutoff so high that significant sensitivity was lost. It is important to note that the interdependence of sensitivity and recall is impossible to avoid: maximum sensitivity cannot be maintained without bringing in large numbers of false positives that negatively impact precision-ML.

To visualize this effect, a confidence class chart (CCC) was generated with each training, as part of our standard metrics. The CCC was used as an alternative visualization of the precision-recall trade-off. Figure 3 shows the CCC for the final epoch of our trained model before clinical laboratory validation was performed. This chart shows, by

		Blastocystis sp	Chilomastix troph	Dientamoeba fragilis	Endolimax amoeba troph	Entamoeba troph nucleus	Entamoeba hartmanni troph	Giardia cyst	Giardia troph	Red blood cells	White blood cells	Yeast
0.05	TP	2205	269	1033	727	405	101	575	191	804	196	1351
	FP	1003	85	230	810	187	91	214	55	654	118	1092
	FN	51	84	285	53	11	27	11	6	51	19	75
0.1	TP	2174	269	1033	715	401	99	569	191	789	194	1323
	FP	798	85	230	626	134	70	174	42	521	96	784
	FN	82	84	285	65	15	29	17	6	66	21	103
0.2	TP	2121	269	1033	693	397	97	556	186	768	185	1272
	FP	620	85	230	422	96	54	147	35	431	76	564
	FN	135	84	285	87	19	31	30	11	87	30	154
0.3	TP	2051	269	1033	666	387	96	538	180	746	179	1221
	FP	517	85	230	339	79	45	129	29	355	63	431
	FN	205	84	285	114	29	32	48	17	109	36	205
0.4	TP	1979	269	1033	635	379	92	521	179	716	177	1147
	FP	421	85	230	270	63	34	115	26	297	57	359
	FN	277	84	285	145	37	36	65	18	139	38	279
0.5	TP	1882	269	998	596	375	89	501	177	681	170	1084
	FP	353	85	201	219	50	31	99	25	259	48	288
	FN	374	84	320	184	41	39	85	20	174	45	342
0.6	TP	1755	269	926	534	365	86	481	176	626	169	1033
	FP	280	85	144	174	42	29	82	21	225	44	239
	FN	501	84	392	246	51	42	105	21	229	46	393
0.7	TP	1582	259	830	480	356	84	455	175	569	166	945
	FP	218	80	101	140	37	26	62	19	186	37	180
	FN	674	94	488	300	60	44	131	22	286	49	481
0.8	TP	1352	247	701	423	343	79	418	169	497	161	827
	FP	150	62	63	105	33	23	46	17	145	33	132
	FN	904	106	617	357	73	49	168	28	358	54	599
0.9	TP	942	219	495	320	320	73	352	164	371	147	652
	FP	71	50	34	64	26	14	35	15	70	27	72
	FN	1314	134	823	460	96	55	234	33	484	68	774
0.95	TP	572	195	316	217	292	70	285	159	252	137	458
	FP	39	43	19	32	22	10	25	12	35	21	36
	FN	1684	158	1002	563	124	58	301	38	603	78	968
0.98	TP	202	160	136	103	251	63	206	152	121	117	241
	FP	9	30	7	8	16	5	14	9	11	15	13
	FN	2054	193	1182	677	165	65	380	45	734	98	1185
0.99	TP	53	129	48	50	223	58	156	146	67	110	125
	FP	2	16	4	1	12	4	10	7	1	9	1
	FN	2203	224	1270	730	193	70	430	51	788	105	1301

FIG 3 The CCC chart shows by parasite class, the number of true positives (TPs), false positives (FPs), and false negatives (FNs) for each class of parasite at increasing confidence score cutoffs. Green indicates TPs (goal: maximize), orange indicates FNs (goal: minimize), and red indicates FPs (goal: minimize). The image was generated using Python Pillow library (<https://python-pillow.org/>).

parasite class, the number of true positives (TPs), false positives (FPs), and false negatives (FNs) for each class of parasite at increasing confidence score cutoffs. Low confidence cutoff yielded maximum TPs (green), but high FPs (red), and minimized FNs (orange), resulting in maximum recall but a low precision-ML. High-confidence cutoff yielded lower TPs (green), minimized FPs (red), and higher FNs (orange), resulting in low recall but high precision-ML. A confidence cutoff of 0.4 was determined to be an acceptable balance of good recall with reasonable precision-ML. This metric was combined with confidence score sorting which strongly biases TPs toward the front of the sequence of images shown to the user. This aggregate performance allowed for predictable success in clinical laboratory validation of the model.

TABLE 2 Contingency table and slide-level agreement calculations comparing the CNN model to the gold-standard O&P examination for 87 true-positive specimens and 106 true-negative specimens^a

CNN model analysis result	O&P examination result (n)	
	Positive	Negative
Positive	86	2
Negative	1	104

^aPositive percent agreement, 98.88% (95% CI, 93.76% to 99.98%); negative percent agreement, 98.11% (95% CI 93.35% to 99.77%).

Slide-level accuracy. Of the 91 slides previously reported as positive for at least one parasite that were scanned and analyzed, 87 (95.6%) had an acceptable scan. Of these 87 slides, 86 were deemed positive after manual analysis of labeled images for a suspect parasite, multiple parasites, or reportable levels of WBCs and RBCs. The resulting slide-level agreement for positive specimens was 98.88% (95% confidence interval [CI], 93.76% to 99.98%). The one slide reported as negative was manually examined by microscopy, and the target organism (*Endolimax nana*) was detected. Of the four slides (4.4%) that had unacceptable scans, three were still unacceptable upon rescan, but the fourth had a successful rescan that yielded agreement with the expected result (1+ *Blastocystis* sp.) (Table 2).

Of the 125 slides previously reported as negative for parasites that were scanned and analyzed, 106 (84.8%) had a successful scan and were analyzed by the model. Of the 125 negative slides, 10 slides (8.0%) failed to scan due to slides containing very little biomass with which to execute initial focus and 9 (7.2%) were reported as unacceptable due to the scans being blurry or having incomplete image analysis. Of these latter nine slides, eight remained unacceptable or failed upon rescan and the ninth had a substantial air bubble in the scan area and could not be reliably scanned. The most common cause of unacceptable and failed scans was an insufficient amount of fecal material on the slides, which was commonly seen with watery diarrheal specimens. Of the 106 acceptable scans, 104 were deemed negative after analysis by the model and subsequent manual image evaluation, for a 98.11% (95% CI, 93.35% to 99.77%) slide-level negative agreement. The two slides that were flagged by the model as containing putative protozoa were manually examined by microscopy, and both slides were confirmed as negative.

Precision-CM analysis. Clinical microbiology precision (precision-CM) studies were performed using the following slides: (i) a slide containing *G. duodenalis*, *E. nana*, and 1+ *Blastocystis*; (ii) a slide containing 1+ *Blastocystis*; and (iii) a negative slide. For the within-run precision studies, all three slides were scanned and analyzed by the model three times on the same day. The model identified the specimens as expected at the slide level. For the between-run precision-CM studies, all three slides were scanned and analyzed by the model once on three different days. The model identified the specimens as expected at the slide level.

Limit of detection. Neat stool and dilutions were tested for 1:1, 1:2, 1:4, 1:8, 1:16, 1:32, 1:64, 1:128, and 1:256. The dilution series were incorporated into the normal O&P workflow and read in a blind manner by technologists with standard O&P processes. For the four series analyzed by the technologists, parasites were not detected below the 1:8 dilution. In all four series, the software detected at least one of the two species at the final 1:256 dilution (Table 3).

DISCUSSION

O&P examinations remain the “go-to” test for protozoal diagnostics in clinical microbiology laboratories despite other readily available (often more sensitive) diagnostic modalities that do not rely on manual microscopy and highly trained morphologists. One limitation of these aforementioned diagnostic modalities is that they are often targeted to only detect one to four pathogenic protozoa. The O&P has historically

TABLE 3 Limit of detection data for the five runs

Dilution	Dilution series 1		Dilution series 2		Dilution series 3		Dilution series 4	
	Technologist	Model	Technologist	Model	Technologist	Model	Technologist	Model
Neat	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	Negative	<i>Giardia duodenalis</i> , 2 + <i>Blastocystis</i> sp.	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.
1:1	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	Negative	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.
1:2	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	Negative	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.
1:4	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	Negative	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.
1:8	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	Negative	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.
1:16	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	Negative	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.
1:32	Negative	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	Negative	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	Negative	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	Negative	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.
1:64	Negative	Negative ^a	Negative	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	Negative	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	Negative	<i>Giardia duodenalis</i>
1:128	Negative	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	Negative	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	Negative	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	Negative	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.
1:256	Negative	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	Negative	<i>Giardia duodenalis</i>	Negative	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.	Negative	<i>Giardia duodenalis</i> , 1 + <i>Blastocystis</i> sp.

^aIncomplete scan.

been viewed as a wider net for which a larger assortment of protozoa may be detected and identified (in spite of many of those protozoa being nonpathogenic). Nonetheless, physician demand for this testing drives a need for further test volume capacity in the clinical laboratory; it may be difficult to sustain and maintain quality and turnaround time in high-volume settings, such as a tertiary care facility or national reference laboratory. The major barrier to growing volume capacity for this testing is the lack of trained personnel and the limitation of how many specimens those trained individuals can accurately evaluate in a given work day without suffering ergonomic and emotional stress as well as poor job satisfaction. Gains in O&P specimen processing efficiency have been realized in some laboratories (3); however, no advancements in the detection step of the O&P have been realized, rendering the process as time-consuming and difficult as it was 50+ years ago. Two aspects that consume the most time for O&P evaluation include the time required to search for putative parasites in a smear before determining that a specimen is truly negative and the time required to detect multiple examples of a suspected parasite when specimens contain low parasite burden. Automated digital scanning of microscope slides and deep machine learning have the potential to augment these processes by providing comprehensive detection of parasites in stool and presenting the findings to a technologist in a user-friendly, condensed pictorial interface. In other words, the manual process of locating suspicious structures in a stained slide can be simplified by artificial intelligence (AI), allowing the technologists to interpret high-resolution images of parasites quickly and comprehensively (Fig. 4).

Standard machine learning metrics (as provided with the ML software used) are challenging to directly apply to this application, as the images being processed are very sparsely populated with true parasites. Consider the following example: a 32,000-pixel by 80,000-pixel scan with 10 parasites. With 50% overlap, the scan will require ~164,000 250-by-250-pixel scenes to process. With 10 parasites on the slide, if we choose to give credit in the metrics for correctly calling “no parasite found” a true background, then 163,990 of 164,000 would be “correct” even if all the model did was report background for every scene. As a result, the data would generate a near-perfect ROC and P-R plot while revealing very little useful information about performance characteristics of the model. Thus, in our posttraining performance evaluation metrics there was no credit (true positive) given for correctly labeling a scene as background (empty). The CCC was a more informative measure of our model’s ability to rule out protozoa with confidence while minimizing false positives.

The primary performance metric of clinical laboratory validation was slide-level accuracy (as defined in Methods and Materials). Operationally, the specimens that were flagged as containing parasites were then queued for a manual microscopic confirmation, resulting in final identification by a trained parasitologist. Negative specimens that were not flagged for any putative parasites and could, thus, be quickly reviewed in the user interface and evaluated via the digital images. Occasionally, the model may have falsely identified suspicious objects with features similar to the established protozoal classes, but for which a manual review of the slide could quickly determine were truly negative.

These study data revealed that digital imaging using the 3DHISTECH Panoramic 250 automated slide scanner when paired with the proprietary AI from Techcyte allowed for excellent slide-level agreement (98.88% [95% CI, 93.76% to 99.98%] positive agreement and 98.11% [95% CI, 93.35% to 99.77%] negative agreement) and precision-CM compared with manual microscopy. The model identified probable parasites in two slides that previously were reported as “negative for parasites” by standard-of-care microscopy. Manual examination of these discrepant specimens confirmed the original result by microscopy. Only a single previously identified positive specimen failed to be detected by the model. The specimen contained rare *E. nana*. Based on the data generated in the LOD studies, the model should identify low-parasite burden specimens more frequently than a manual interrogation. In these studies, the model was at least 5 serial dilutions more sensitive (analytically) for the detection of *G.*

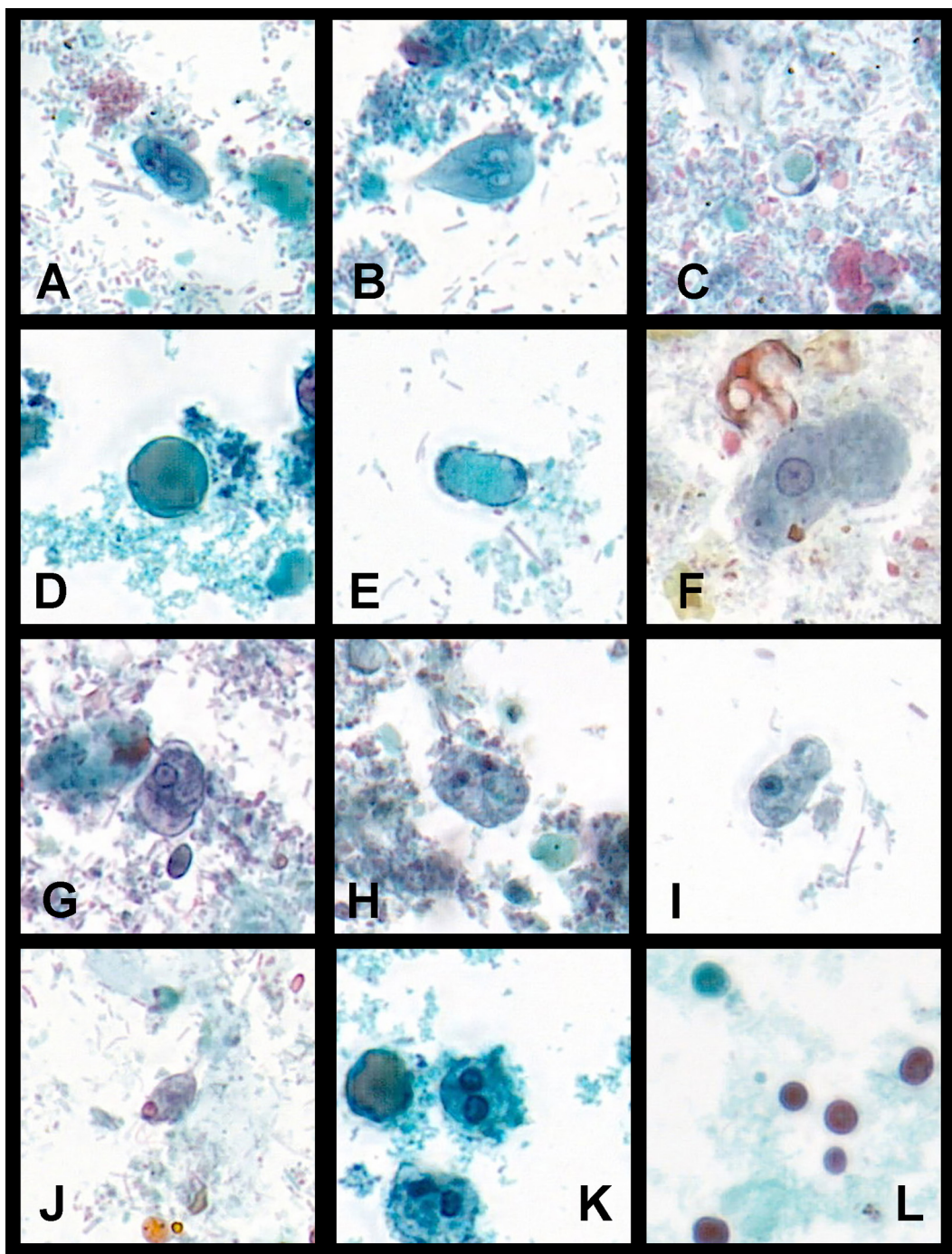


FIG 4 Representative examples of each unique class detected by the model and presented in the Techcyte software for analysis. (A) *Giardia duodenalis*, cyst. (B) *Giardia duodenalis*, trophozoite. (C) *Blastocystis* sp., cyst-like form. (D) *Blastocystis* sp., vacuolar form. (E) *Blastocystis* sp., dividing forms. (F) *Entamoeba* sp., non-hartmanni, trophozoite. (G) *Entamoeba hartmanni*, trophozoite. (H) *Dientamoeba fragilis*. (I) *Endolimax nana*, trophozoite. (J) *Chilomastix mesnili*, trophozoite. (K) White blood cells. (L) Red blood cells.

duodenalis and *Blastocystis* sp. than was a standard-of-care human-evaluated O&P. In one instance, the technologist failed to identify any protozoa in the neat specimen or LOD dilutions, which is likely attributable to having less than 1 month of practical experience. This phenomenon is likely also relevant to laboratories with low O&P volume and low prevalence of positives, where staff may have difficulty gaining and maintaining competence. These aggregate performance characteristics allow for con-

fidant integration of the model and scanner into routine clinical care for augmented O&P workflow purposes. Adoption of this tool should provide equal or greater diagnostic yield than that of a human performing microscopic examinations, while also providing a user-friendly process for specimen evaluation.

The model can be integrated into an existing clinical parasitology laboratory in an effort to rapidly screen out negative specimens with high confidence. Images of putative parasites in a truly negative slide may occasionally be presented to the technologist in the digital interface, but true negative specimens typically only contain a few images that require evaluation. If the images appear false, the technologist can document that the slide is negative and proceed to the next specimen. If the images are suspicious, the slide can be manually examined to confirm or deny the findings of the model. When images are captured from a scan that represent true parasites, the technologist will manually confirm and report accordingly. In this sense, the software is augmenting the standard process, while not determining the ultimate identification. In total, an individual slide requires 4 to 5 minutes of scan time and 0.5 to 3 minutes of technologist evaluation time. While class confusion (e.g., detecting *Giardia* sp. that is actually *Dientamoeba* sp.) may decrease as the model improves, early integration in clinical care (considering the low positivity) should still warrant a manual evaluation for suspicious slides. One important point to consider is that the model could be affected by multiple variables in slide preparation and imaging. The specific model could perform differently in another laboratory setting due to differences in the trichrome stain reagents, smearing technique, and scanner used for data acquisition. In this regard, the model itself is not necessarily “plug-and-play” and would require standardization of preanalytical processes.

This study has several limitations. First, due to the relative scarcity of specimens that contained certain classes, a true hold-out set for evaluation and accuracy of the CNN model, prior to clinical laboratory validation was not employed. This was intentionally designed in an attempt to propagate the model with a greater diversity of class exemplars from multiple unique patient slides, rather than propagating with more exemplars from fewer unique patient slides at the expense of maintaining a unique holdout set. This was an important concession, as training with a greater diversity of exemplars should result in a more robust model in terms of the comprehensive detection of morphologically variable protozoa within a single class. In order to validate the CNN model for clinical laboratory use, an additional set of slides was also required, which the model cannot have previously analyzed. This further constrained the number of unique/rare positive specimens available for holdout testing. Second, we employed an unconventional approach to determine positive and negative agreement by considering the model as part of an AI-augmented O&P rather than a standalone assay. If traditional performance characteristics for agreement were calculated, they would grossly misrepresent the utility of the augmentation (see supplemental methods and Table S1).

Areas of ongoing research in this immediate application of the AI technology include training the existing model to identify the current classes with better organism-level accuracy to improve the ease of manual interpretation of the labeled organisms. Excellent class-level (e.g., organism-level) accuracy should be possible with a much larger and diverse subset of training slides. Due to the rarity of these organisms, a collection of additional slides will occur prospectively over time in order to enrich the data set and refine the model. Additionally, new classes, such as *Cyclospora*, *Pentatrichomonas*, and *Enteromonas*, will be trained on the trichrome stain model in an effort to improve the use of this method as an augmentation of the manual examination. Modified acid-fast stains are also an area of investigation that can complement this model by allowing the rapid detection of *Cyclospora* sp. and *Cryptosporidium* sp. Work-flow analysis studies are also underway to quantitatively evaluate the impact of adopting this AI in a high-volume parasitology laboratory.

This work adds to the already growing literature showing the value of AI in different fields of medicine, such as oncology, radiology, and most recently for areas of bacteriology related to gram stains and antibiotic resistance in phenotypic growth assays

(17–21). To our knowledge, this is the first description of a CNN used to augment the detection of protozoa in human clinical O&P specimens stained with trichrome. This augmentation provides slide-level accuracy equal to a human evaluation, with superior analytical sensitivity.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 0.6 MB.

ACKNOWLEDGMENTS

We thank the parasitology laboratory at ARUP laboratories for their assistance in laboratory validation steps and specimen collection. We thank Pramika Stephan for early proof-of-concept scanner evaluation work. We also appreciate the collaborative efforts of Techcye, Inc., staff, including Ben Cahoon, Ben Martin, and Joseph Szendre.

This work could not have been made possible without the financial support of the ARUP Institute for Clinical and Experimental Pathology, supported generously by Mark Astill and Adam Barker.

R.B.S. and J.F.W. are employed by and hold stock ownership in Techcye, Inc.

O.A.'s primary contributions included early project conceptualization, proof-of-concept testing, technology evaluation, and direction of model development/data acquisition. M.R.C.'s primary contributions included scientific and clinical direction of model development/data acquisition and assay validation.

REFERENCES

- Garcia LS, Arrowood M, Kokoskin E, Paltridge GP, Pillai DR, Procop GW, Ryan N, Shimizu RY, Visvesvara G. 2018. Laboratory diagnosis of parasites from the gastrointestinal tract. *Clin Microbiol Rev* 31:e00025-17. <https://doi.org/10.1128/CMR.00025-17>.
- Garcia E, Kundu I, Ali A, Soles R. 2018. The American Society for Clinical Pathology's 2016–2017 vacancy survey of medical laboratories in the United States. *Am J Clin Pathol* 149:387–400. <https://doi.org/10.1093/ajcp/aqy005>.
- Couturier BA, Jensen R, Arias N, Heffron M, Gubler E, Case K, Gowans J, Couturier MR. 2015. Clinical and analytical evaluation of a single-vial stool collection device with formalin-free fixative for improved processing and comprehensive detection of gastrointestinal parasites. *J Clin Microbiol* 53:2539–2548. <https://doi.org/10.1128/JCM.00838-15>.
- George E. 2010. Occupational hazard for pathologists: microscope use and musculoskeletal disorders. *Am J Clin Pathol* 133:543–548. <https://doi.org/10.1309/AJCPUXD5SKJRFVW>.
- Gopakumar GP, Swetha M, Sai Siva G, Sai Subrahmanyam G. 2018. Convolutional neural network-based malaria diagnosis from focus stack of blood smear images acquired using custom-built slide scanner. *J Biophotonics* 11:e201700003. <https://doi.org/10.1002/jbio.201700003>.
- Rosado L, da Costa JMC, Elias D, Cardoso JS. 2017. Mobile-based analysis of malaria-infected thin blood smears: automated species and life cycle stage determination. *Sensors (Basel)* 17:2167. <https://doi.org/10.3390/s17102167>.
- Diaz G, Gonzalez FA, Romero E. 2009. A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images. *J Biomed Inform* 42:296–307. <https://doi.org/10.1016/j.jbi.2008.11.005>.
- Holmstrom O, Linder N, Ngasala B, Martensson A, Linder E, Lundin M, Moilanen H, Suutala A, Diwan V, Lundin J. 2017. Point-of-care mobile digital microscopy and deep learning for the detection of soil-transmitted helminths and *Schistosoma haematobium*. *Glob Health Action* 10:1337325. <https://doi.org/10.1080/16549716.2017.1337325>.
- Intra J, Taverna E, Sala MR, Falbo R, Cappellini F, Brambilla P. 2016. Detection of intestinal parasites by use of the cuvette-based automated microscopy analyser sediMAX(R). *Clin Microbiol Infect* 22:279–284. <https://doi.org/10.1016/j.cmi.2015.11.014>.
- Nkamgang OT, Tchiotsop D, Tchinda BS, Fotsin HB. 2018. A neuro-fuzzy system for automated detection and classification of human intestinal parasites. *Inform Med Unlocked* 13:81–91. <https://doi.org/10.1016/j.imu.2018.10.007>.
- Tchinda BH, Tchiotsop D, Tchinda R, Wolf D, Noubom M. 2015. Automatic recognition of human parasitic cysts on microscopic stools images using principal component analysis and probabilistic neural network. *Ijarai* 4:26–33. <https://doi.org/10.14569/IJARAI.2015.040906>.
- Yang YS, Park DK, Kim HC, Choi MH, Chai JY. 2001. Automatic identification of human helminth eggs on microscopic fecal specimens using digital image processing and an artificial neural network. *IEEE Trans Biomed Eng* 48:718–730. <https://doi.org/10.1109/10.923789>.
- Ghazali KH, Hadi RS, Mohamed Z. 2013. Automated system for diagnosis intestinal parasites by computerized image analysis. *Mod Appl Sci* 7:98–114. <https://doi.org/10.5539/mas.v7n5p98>.
- Nesterov Y. 1983. A method of solving a convex programming problem with convergence rate $O(1/k^2)$. *Soviet Math Dokl* 27:372–376.
- de Boer P-T, Kroes D, Reuven S, Rubinstein RY. 2005. A tutorial on the cross-entropy method. *Ann Oper Res* 134:19–67. <https://doi.org/10.1007/s10479-005-5724-z>.
- Davis J, Goadrich M. The relationship between precision-recall and ROC curves. 2006. In *Proceedings of the 23rd International Conference on Machine Learning*, Pittsburgh, PA.
- Smith KP, Richmond DL, Brennan-Kohn T, Elliott HL, Kirby JE. 2017. Development of MAST: a microscopy-based antimicrobial susceptibility testing platform. *SLAS Technol* 22:662–674. <https://doi.org/10.1177/2472630317727721>.
- Smith KP, Kang AD, Kirby JE. 2017. Automated interpretation of blood culture gram stains by use of a deep convolutional neural network. *J Clin Microbiol* 56:e01527-17. <https://doi.org/10.1128/JCM.01521-17>.
- Brinker TJ, Hekler A, Enk AH, Berking C, Haferkamp S, Hauschild A, Weichenthal M, Klode J, Schadendorf D, Holland-Letz T, von Kalle C, Fröhling S, Schilling B, Utikal JS. 2019. Deep neural networks are superior to dermatologists in melanoma image classification. *Eur J Cancer* 119:11–17. <https://doi.org/10.1016/j.ejca.2019.05.023>.
- Madani A, Arnaout R, Mofrad M, Arnaout R. 2018. Fast and accurate view classification of echocardiograms using deep learning. *NPJ Digit Med* 1:6. <https://doi.org/10.1038/s41746-017-0013-1>.
- Litjens G, Sanchez CI, Timofeeva N, Hermens M, Nagtegaal I, Kovacs I, Hulsbergen-van de Kaa C, Bult P, van Ginneken B, van der Laak J. 2016. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Sci Rep* 6:26286. <https://doi.org/10.1038/srep26286>.