Use of digital pathology and artificial intelligence for the diagnosis of Helicobacter pylori in gastric biopsies

Daniel S. Liscia¹, Mariangela D’Andrea¹, Elena Biletta¹, Donata Bellis¹, Kejsi Demo¹, Franco Ferrero², Alberto Petti³, Roberto Butinar⁴, Enzo D’Andrea⁴, Giuditta Davini⁴

¹ Unit of Pathology, Department of Surgery ASL BI, Nuovo Ospedale degli Infermi, Ponderano (BI) Italy; ² Unit of Gastroenterology, Department of Surgery ASL BI, Nuovo Ospedale degli Infermi, Ponderano (BI) Italy; ³ Unit of Clinical Engineering ASL BI, Nuovo Ospedale degli Infermi, Ponderano (BI) Italy; ⁴ Engineering 365 Srl, Lancenigo di Villorba (TV) Italy

Summary

Objective. A common source of concern about digital pathology (DP) is that limited resolution could be a reason for an increased risk of malpractice. A frequent question being raised about this technology is whether it can be used to reliably detect Helicobacter pylori (HP) in gastric biopsies, which can be a significant burden in routine work. The main goal of this work is to show that a reliable diagnosis of HP infection can be made by DP even at low magnification. The secondary goal is to demonstrate that artificial intelligence (AI) algorithms can diagnose HP infections on virtual slides with sufficient accuracy.

Methods. The method we propose is based on the Warthin-Starry (W-S) silver stain which allows faster detection of HP in virtual slides. A software tool, based on regular expressions, performed a specific search to select 679 biopsies on which a W-S stain was done. From this dataset 185 virtual slides were selected to be assessed by WSI and compared with microscopy slide readings. To determine whether HP infections could be accurately diagnosed with machine learning, AI was used as a service (AIaaS) on a neural network-based web platform trained with 468 images. A test dataset of 210 images was used to assess the classifier performance.

Results. In 185 gastric biopsies read with DP we recorded only 4 false positives and 4 false negatives with an overall agreement of 95.6%. Compared with microscopy, defined as the “gold standard” for the diagnosis of HP infections, WSI had a sensitivity and specificity of 0.95 and 0.96, respectively. The ROC curve of our AI classifier generated on a testing dataset of 210 images had an AUC of 0.938.

Conclusions. This study demonstrates that DP and AI can be used to reliably identify HP at 20X resolution.

Key words: digital pathology, artificial intelligence, helicobacter pylori

Introduction

With the advent of the COVID-19 pandemic many pathology departments are considering the benefits of remote reporting by DP. Although previous validation studies support the safe use of DP in primary gastrointestinal pathology, a common source of concern about whole slide imaging (WSI) is that limited image resolution could be a reason for an increased risk of malpractice. In a systematic analysis 10% discordance was attributed to the inability to find a small diagnostic/prognostic object on WSI. A lack of image clarity is known to be associated with difficulties in the identification of microorganisms, a commonly cited challenge encountered in DP reporting. Indeed,
a frequent question raised about this technology is whether *Helicobacter pylori* (HP) can be reliably de-
tected by WSI in gastric biopsies, which may consti-
tute a considerable proportion of a pathologist’s daily
workload. In a study conducted by Snead et al. ⁶, pathologists felt that WSI was directly responsible
for a variance in reporting HP and slides needed a
re-scan at a higher magnification to become visible.
While our study was under review, Mayall et al. pub-
lished their findings on multisite networked digital
pathology reporting in England. In this study, seven
out of nine significant diagnostic discrepancies con-
cerned misrecognition of HP in gastric biopsies ⁷. HP
is a gram-negative, spiral-shaped bacterium that in-
fests 50% of the world’s population ⁸ and is known to
be a major cause of chronic gastritis, peptic ulcers,
gastric adenocarcinomas and lymphomas ⁹. In 1994,
the World Health Organization (WHO) and the Inter-
national Agency for Research on Cancer classified
HP as a class 1 carcinogen ¹⁰. Although several non-invasive methods for the de-
tection of HP infection are available ¹¹, this diagnosis
often is made by conventional histology on gastric bi-
opsies taken because of endoscopic abnormalities ¹². Diag-
nosis of HP can be performed in hematoxylin and
eosin (H&E) stained sections with a reported high
specificity only when moderate to severe lymphocytic
infiltration are present. However, H&E diagnosis of HP
can be tedious and time consuming with a sensitiv-
ity that drops to 3% when plasmacytic, lymphocytic,
and neutrophilic infiltration are absent or mild ¹³. Fur-
thermore inter-observer variability achieved on H&E
diagnosis was found to be high and unacceptably sub-
jective ¹⁴. This can be improved by special stains such as Gi-
emsa, Warthin-Starry silver (W-S), Genta, and immu-

![Figure 1](image1.png)

**Figure 1.** Illustrative figures of a HP positive case (3+) displayed by 20X digital images. A) H&E stain B) organisms are clearly
visible by W-S stain. The following figures (C and D) are an example of a false positive 1+ score. This case was in the AI testing
set and shows pseudo-bacterial particles mistaken for HP by the algorithm.
nohistochemistry. Even though immunohistochemistry is more sensitive and more specific, it is not an up-front ancillary study. Giemsa stain is the most commonly used in first-line routine clinical practice. However, with this histochemical method, a high resolution (40X) digital scan is required to maintain the diagnostic performances of traditional microscopy. The main goal of this work is to show that a reliable diagnosis of HP infection can be made by DP even at relatively low magnifications. A high viewing magnification ($\times40$, 0.23 μm/pixel) requires large amounts of storage space and bandwidth consumption, affecting the speed of image update.

The method we propose is based on the W-S stain which enhances the visualization of HP and allows its faster and more accurate detection (Fig. 1). To check for any differences, diagnoses of HP infection made with WSI were compared with those made with traditional microscopy (TM). The secondary goal is to demonstrate that an AI algorithm, based on “Deep Learning” can diagnose HP infections with sufficient accuracy on DP images. Large datasets generated by DP have offered additional opportunities for AI-aided diagnostics which can classify slides based on their specific patterns. This approach can be used only after it has been established, with a high degree of certainty, that the diagnostic performances of TM and WSI are comparable to the extent that one can replace the other. As is well known, HP detection in histological slides is a repetitive task that comprises a significant part of a pathologist’s workload. If an AI algorithm could reach a point to be able to render automatic this part of diagnostic work, substantial benefits could be obtained.

As already done by Klein et al., we used a convolutional neural network (CNN) based deep learning system and similarly the training set was made of smaller image patches of subdivided whole slide images. But differently from the cited study, we used 20X scanned slides for the detection of HP. In addition in the present study, AI was used as a Service (AIaaS) from a third-party that provides out-of-the-box AI solutions, a novel approach that benefits from the availability of deep learning (DL) web platforms without cost-prohibitive investments in massive hardware and programmers.

Materials and methods

The histopathology laboratory is equipped with a high resolution scanner (Nanozoomer-XR, Hamamatsu Photonics K.K., Japan) that is capable of rapid automatic processing of up to 320 slides. As previously described a web platform for remote access to virtual slide trays (Cloud Pathology Group, Milan, Italy) allowed case searching by unique case numbers. All slides in this study were anonymized and scanned at 20X (0.46 μm/pixel) resolution to save archiving space and to speed up image download. The scanner is connected to the hospital network and its software receives slides data from the anatomic pathology laboratory information system (LIS) (Winsap, Turin, Italy).

Case selection

We retrospectively reviewed pathology reports of gastric biopsies made between January 2019 and September 2020 and stored them in our LIS database. All biopsies were performed both in the antrum and in the corpus with at least two samples for each region according to European guidelines. Through a query 1783 cases of gastric biopsies were retrieved and saved as comma separated values (csv) file containing the accession number and the complete text of the final diagnosis. Within this data set, we selected all biopsies on which a W-S stain was done so that they could be searched in our virtual slide repository.

To accomplish this, a more specific search was conducted by designing a software tool developed in the Python programming language (Python 3.7) that used regular expressions to perform pattern matching in pathology reports. With this software unstructured free-text could be turned into computationally usable data and stored in tabular format (Python sources available in Supplemental File). The regular expression tool first searched for all diagnoses in which HP was looked for, then on those cases a search was done to select only those biopsies in which the “Warthin-Starry” substring occurs at least once. On these latter cases a third level search was performed to separate HP positive from HP negative biopsies. Finally, only in HP positive biopsy reports the strings “1+”, “2+” and “3+” were searched as HP were classified into three grades, in accordance with the Updated Sydney System reflecting the average density on the surface and the foveolar epithelium. Using this search tool, 679 biopsies were identified and, to approximate actual daily working conditions, from this set each pathologist examined 46 slides. As a result 185 virtual slides were selected to be assessed by WSI and compared with microscopy slide readings. All pathologists were allowed to view the corresponding digital slides of H&E stained sections of the cases on which a Warthin-Starry stain was available. To maintain a well-balanced ratio between cases, 101 biopsies were HP negative and 84 were HP positive. Digital slides were read by four independent expert pathologists.
with more than 8 years experience, blind to the TM diagnosis, defined as reference standard. To establish a strong gold standard and also to determine what types of slides can lead to discrepancies, discordant cases were reviewed for a consensus from all pathologists.

**Warthin-Starry Staining**

W-S silver stain was performed using the Artisan automatic stainer (Artisan Link Pro Special Staining System, Agilent Dako, Denmark), capable of reproducibly process up to 48 slides in around 7 minutes. The procedure was performed following manufacturer’s instructions, as previously published.

**Artificial Intelligence**

To test whether HP infections could accurately be diagnosed by DL we used an AI as a service (AaaS) web platform optimized to analyze images of no more than 4 megapixels (Microsoft Custom Vision, Redmond, Washington, USA). Based on their visual characteristics a classification model was created that learned from a series of labeled histological fields of W-S stained biopsies.

In this study, the recognition of HP by the classifier was evaluated independently from the gastric inflammatory status. For the training and testing of the predictive model we used two subsets of cases in which there was an absolute subclass concordance between the TM and WSI diagnoses.

From a set of HP positive and HP negative WSI cases, two distinct groups (Fig. 2) were randomly selected:

1. the first group served as a labeled training set for a supervised learning of the classifier. It was composed of 12 virtual slides at 20X magnification. From this group, a series of images with an average size of 2000×2000 pixels were cropped, 229 representing HP positive fields and 239 HP negative fields. This training set was initially used to build the model.

2. the second group, used for testing, was composed of 69 HP positive and 141 HP negative cropped images. The performance of the classifier was evaluated using this testing set.

This part of the image cropping and file processing was automatically managed by a python script (Fig. 3).

**Statistical analysis**

Most data processing, plotting and statistical analysis were performed using Python 3.7 and the related packages (Pandas, NumPy, SciPy). Cramer’s Phi test and the Kendall’s tau test were uses as a measure of correlation. Cohen’s kappa statistical coefficient was used to test inter-rater agreement of categorical data when WSI and TM were considered as two distinct
raters, rating the same case. A ROC curve and the corresponding area under the curve (AUC) was plotted to evaluate the diagnostic performance of the machine learning classification model at all thresholds. The performance measurement for the effectiveness of our supervised ML classifier was achieved by computing a confusion matrix \[\text{30}\], accuracy, precision, recall (sensitivity) and the F1 score, a weighted average of the precision and recall.

This study was conducted following approval of the Institutional Review Board of the ASL BI, Nuovo Ospedale degli Infermi (study n. CE 291/20).

**Results**

**Digital Pathology**

Of 185 cases reviewed over WSI, we recorded 4 false positives and 4 false negatives with an overall agreement of 95.6\%. Table I shows the significant association between TM and WSI (\(p < 0.0001\)) and both modalities detected 101 HP negative and 84 HP positive cases, with an excellent correlation (Pearson's Phi = 0.913).

If we looked at what was the HP score assigned to the false positive and false negative cases, we found...
that both fell into the “1+” category. Compared with TM, defined as the “gold standard” for the diagnosis of HP infections, WSI had a sensitivity, specificity and accuracy of 0.95, 0.96 and 0.95, respectively (Tab. II). However when the semi-quantitative assessment of HP infection was rated on the three point scale (1+ to 3+) the correlation, although significant (p < 0.001), was rather low (Kendall’s tau = 0.594). If these two reading methods were considered as two different observers and a Cohen’s kappa for inter-observer agreement was computed, we obtained a value of 0.137, while if only the positive and negative were compared, the value was close to 1 (Cohen’s kappa = 0.913).

**ARTIFICIAL INTELLIGENCE**

The confusion matrix (Tab. III) describes the performance of our classification model on the set of 210 test images for which the HP values were known.

**Table III.** The Confusion Matrix describes the performance of our supervised machine learning classification model on a set 210 test images cropped from DP virtual slides for which the true HP diagnoses were known.

| Predicted HP negative | Predicted HP positive | Totals |
|-----------------------|-----------------------|--------|
| Actual HP negative    | 123                   | 18     | 141    |
| Actual HP positive    | 7                     | 62     | 69     |
| Totals                | 130                   | 80     | 210    |

Chi²(4, N = 210) = 116.74, p < 0.0001; Pearson’s Phi Coefficient = 0.75.

**Table IV.** Measures of performance of our deep learning classifier that provide an understanding of the diagnostic accuracy of the trained model. The F1 score, the harmonic mean of the precision and recall, is a measure of the classifier selecting power.

|                     |                        |
|---------------------|------------------------|
| Accuracy            | 0.880                  |
| True positive rate  | 0.997                  |
| True negative rate  | 0.872                  |
| Positive Predictive Value | 0.772          |
| Negative Predictive Value | 0.946          |
| F1 score            | 0.829                  |

Figure 4. ROC Curve showing true positive rate vs. false positive rate for the detection of HP at different thresholds of our ML classifier. The diagonal red line serves as a reference since it is the curve of a classifier that selects cases at random.

Among the image fields of the testing data-set the true positives were 62, the true negative 123, the false positive and false negatives were 18 and 7, respectively (p < 0.0001). Table IV recapitulates the diagnostic performance of our DL algorithm. The balance accuracy and F1 score values of 0.88 and 0.82, respectively, are metrics of the binary classifier, particularly useful when the classes are imbalanced as in our case, taking in account the higher frequency of HP negative cases compared to the HP positive. To gauge the diagnostic power of the ML approach, a ROC curve was plotted (Fig. 4) that gave a AUC of 0.93. Our machine learning model was also tested against the grade of HP infection giving, not surprisingly, only a mediocre result in the relationship between the actual and the predicted ranks (Kendall Tau = 0.58).
Discussion

In the first part of our study, we investigated whether WSI can be used to reliably diagnose HP infections in gastric biopsies. This was not only our concern, since the same doubts have been raised at numerous meetings we have attended and in several articles on DP. To facilitate the diagnosis of HP infections with WSI we have adopted W-S silver stain, a method that allows faster identification of HP even at a low resolution.

One of the aspects that may be stressed of the present work is methodological. All the cases needed for the WSI, and AI were selected using a software tool expressly designed to extract structured data from unstructured textual reports. Pathologists were allowed to see the H&E stained slides, and therefore they were aware of the gastric inflammatory state. This, of course, facilitated the pathologist’s task to render a proper diagnosis for the known relationship between HP and gastritis.

Only 8 cases (4.3%) read with the WSI did not correspond to the initial diagnosis since 4 false positives and 4 false negatives were found. The latter of course occurs in practice, but the former, relatively rare, can result from a bacterial overgrowth due to proton pump inhibitors. All the WSI diagnoses discordant with those carried out under the microscope fell into the category of “1+”, the one in which the degree of diagnostic agreement between pathologists is rather low. WSI was absolutely comparable with TM in its ability to discriminate between HP negative and HP positive gastric biopsies, as demonstrated by the high level of accuracy. The comparison between WSI and TM was also analyzed from a different point of view by considering the computer and the microscope as two different observers. The kappa statistic indicates an almost perfect agreement between the two modalities. Much lower was the agreement in assessing the degree of HP infection using a three tiers rank, showing that in routine pathology an absolute rigor in grading the HP density is by no means a trivial exercise. However, for clinical management the most important piece of information is whether HP is present. The data presented in our study clearly demonstrate that using the W-S stain a diagnosis of HP infections with WSI is both possible and accurate even at a 20X magnification.

These results convinced us that WSI could be used to train a DL algorithm with an AI based image recognition web platform. By examining the DL Confusion Matrix, if we compared the actual and the predicted HP status, our classifier was able to make a correct diagnosis in over 80% of cases. The level of discordance is apparently high, but it must be stressed that, contrary to the pathologists, the ML classifier was at a disadvantage not being informed of the degree and type of gastritis. Despite this, the ROC analysis showed an AUC of 0.938 and demonstrated the accuracy of our classifier.

We are convinced that if the algorithm could directly access, as a training set, all the digital archive of W-S stained biopsies together with the H&E virtual slides, the classifier would exceed the accuracy of a pathologist. This scenario has been the subject of interesting studies recently published by Campanella et al. and by Iizuka et al. As noted by the latter author, the main challenge is the sheer size of a WSI that can contain several billion pixels, while the area of interest can be extremely limited. We show that we are at a point in which AI is no longer optional or limited to a small group of experts. We believe that our approach to AI will broaden its applicability to the diverse and challenging problems of histopathology image analysis.

Conclusions

We believe that the method used in this work accomplished the main goal of reliably diagnose HP infections with WSI even at a 20X magnification. The secondary goal was also met, as we have shown that AI algorithms, based on deep learning, can diagnose HP infections with sufficient accuracy on DP images.

One of the aspects that may be stressed in the present study is methodological. All the cases needed for WSI and AI were selected using a software tool expressly designed to extract structured data from unstructured textual reports. Another methodological aspect is the use of AI as a service (AiaaS) a ready-made AI solution delivered over a web platform for the automatic recognition of histological images. To our knowledge, this has never been done before with WSI.

ACKNOWLEDGMENTS

We wish to thank the Fondazione Edo ed Elvo Tempia, Biella, Italy, and the Fondazione Cassa di Risparmio di Biella for their support.

We wish also to thank Prof. Mark Bernheim for his thoughtful suggestions and Prof. Umberto Dianzani for his invaluable help and guidance.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.
**Funding**

This study was financially supported by Fondazione Edo ed Elvo Tempia, Biella, Italy, and by Fondazione Cassa di Risparmio di Biella.

**References**

3 Williams BJ, Brettie D, Aslam M, et al. Guidance for Remote Reporting of Digital Pathology Slides During Periods of Exceptional Service Pressure: An Emergency Response from the UK Royal College of Pathologists. J Pathol Inform 2020;11:12. https://doi.org/10.4103/ipi.ipi_23_20

4 Loughrey MB, Kelly PJ, Houghton OP, et al. Digital slide viewing for primary reporting in gastrointestinal pathology: a validation study. Virchows Arch 2015;467:137-144. https://doi.org/10.1007/s00428-015-1780-1

5 Williams BJ, DaCosta P, Goacher E, et al. A Systematic Analysis of Discordant Diagnoses in Digital Pathology Compared With Light Microscopy. Arch Pathol Lab Med 2017;141:1712-1718. https://doi.org/10.1043/arp.a.2016-0494-OA

6 Araújo ALD, Arboleda LPA, Palmier NR, et al. The performance of digital microscopy for primary diagnosis in human pathology: a systematic review. Virchows Arch 2019;474:269-287. https://doi.org/10.1007/s00428-018-02519-z

7 Lujan GM, Savage J, Shana’ah A, et al. Digital pathology initiatives and experience of a large academic institution during the Coronavirus disease 2019 (COVID-19) pandemic. Archives of Pathology & Laboratory Medicine 2021. https://doi.org/10.5858/arpa.2020-0715-SA

8 Snead DRJ, Tsang Y-W, Meskiri A, et al. Validation of digital pathology imaging for primary histopathological diagnosis. Histopathology 2016;68:1063-1072. https://doi.org/10.1111/his.12879

9 Mayall FG, Smethurst H-B, Semkin L, et al. A feasibility study of multisite networked digital pathology reporting in England. Journal of Pathology Informatics 2022;13:4. https://doi.org/10.4103/ipi.ipi_61_21

10 Marshall BJ, Warren JR. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. Lancet 1984;1:1311-1315. https://doi.org/10.1016/S0140-6736(84)91816-6

11 Sipponen P, Price AB. The Sydney System for classification of gastritis 20 years ago. J Gastroenterol Hepatol 2011;26 Suppl 1:31-34. https://doi.org/10.1111/j.1440-1746.2010.06536.x

12 Wu J-Y, Lee Y-C, Graham DY. Eradication of Helicobacter pylori for primary reporting in gastrointestinal pathology: an overview. Diagnostic Histopathology 2019;25:92-99. https://doi.org/10.1055/a-0859-1883

13 Farouk WI, Hassan NH, Ismail TR, et al. Warthin-Starry Staining for primary reporting in gastrointestinal pathology: an overview. Am J Clin Pathol 2018;149:181-185. https://doi.org/10.1093/ajcp/aqx154

14 Garcia-Rojo M. International Clinical guidelines for the adoption of digital pathology: a review of technical aspects. Pathobiology 2016;83:99-109. https://doi.org/10.1159/000441192

15 Moxley-Wyles B, Colling R, Verrelli C. Artificial intelligence in pathology: an overview. Diagnostic Histopathology 2020;26:513-520. https://doi.org/10.1016/j.dmpdh.2020.08.004

16 Tizhoosh HR, Pantanowitz L. Artificial intelligence and digital pathology: Challenges and opportunities. Journal of Pathology Informatics 2018;9:38. https://doi.org/10.4103/ipi.ipi_53_18

17 Klein S, Gildenblatt J, Ihle MA, et al. Deep learning for sensitive detection of Helicobacter pylori in gastric biopsies. BMC Gastroenterology 2020;20:417. https://doi.org/10.1186/s12876-020-01494-7

18 King T. The 16 Best Data Science and Machine Learning Platforms for 2021. Best Business Intelligence and Data Analytics Tools, Software, Solutions & Vendors 2021. Available at: https://solutionsreview.com/business-intelligence/the-best-data-science-and-machine-learning-platforms/. Accessed February 5, 2021.

19 Liscia DS, Bellsis D, Biletta E, et al. Whole-slide imaging allows pathologists to work remotely in regions with severe logistical constraints due to Covid-19 pandemic. J Pathol Inform 2020;11:20. https://doi.org/10.4103/ipi.ipi_32_20

20 Pimentel-Nunes P, Libânio D, Marcos-Pinto R, et al. Management of epithelial precancerous conditions and lesions in the stomach (MAPS II): European Society of Gastrointestinal Endoscopy (ESGE), European Helicobacter and Microbiota Study Group (EHMSG), European Society of Pathology (ESP), and Sociedade Portuguesa de Endoscopia Digestiva (SPED) guideline update 2019. Endoscopy 2019;51:365-388. https://doi.org/10.1055/a-0859-1883

21 G. van Rossum (Guido). Python reference manual. Department of Computer Science [CS] 1995. Available at: https://www.python.org/ pub/5008. Accessed December 27, 2020.

22 Thompson K. Programming Techniques: Regular expression search algorithm. Commun. ACM 1968;11:419-422. https://doi.org/10.1145/363347.363387

23 Farouk WI, Hassan NH, Ismail TR, et al. Classification and grading of gastritis. The updated Sydney System. Internation Workshop on the Histopathology of Gastritis, Houston 1994. Am J Surg Pathol 1996;20:1161-1181. https://doi.org/10.1097/00000478-199610000-00001

24 Chen XY, van der Hulst RW, Bruno MJ, et al. Interobserver variation in the histopathological scoring of Helicobacter pylori related gastritis. J Clin Pathol 1999;52:612-615.

25 Warthin AS, Stary AC. The Staining of Spirochetes in Cover-Glass Smears by the Silver-Agar Method. The Journal of Infectious Diseases 1922;30:592-600.

26 Farouk WI, Hassan NH, Ismail TR, et al. Warthin-Starry Staining for the Detection of Helicobacter pylori in Gastric Biopsies. Malays J Med Sci 2018;25:92-99. https://doi.org/10.21315/mjms2018.25.4.9

27 King T. The 16 Best Data Science and Machine Learning Platforms for 2021. Best Business Intelligence and Data Analytics Tools, Software, Solutions & Vendors 2021. Available at: https://solutionsreview.com/business-intelligence/the-best-data-science-and-machine-learning-platforms/. Accessed February 5, 2021.

28 Tizhoosh HR, Pantanowitz L. Artificial intelligence and digital pathology: Challenges and opportunities. Journal of Pathology Informatics 2018;9:38. https://doi.org/10.4103/ipi.ipi_53_18

29 Chen XY, van der Hulst RW, Bruno MJ, et al. Interobserver variation in the histopathological scoring of Helicobacter pylori related gastritis. J Clin Pathol 1999;52:612-615.

30 Warthin AS, Stary AC. The Staining of Spirochetes in Cover-Glass Smears by the Silver-Agar Method. The Journal of Infectious Diseases 1922;30:592-600.

31 Farouk WI, Hassan NH, Ismail TR, et al. Warthin-Starry Staining for the Detection of Helicobacter pylori in Gastric Biopsies. Malays J Med Sci 2018;25:92-99. https://doi.org/10.21315/mjms2018.25.4.9

32 Ting KM. Confusion Matrix. In: Sammut C, Webb GI, eds. Encyclopedia of Machine Learning. Boston, MA: Springer US; 2010:209-209. https://doi.org/10.1007/978-0-387-30164-8_157

33 Norgan AP, Suman VJ, Brown CL, et al. Comparison of a Medical-Grade Monitor vs Commercial Off-the-Shelf Display for Mitotic Figure Enumeration and Small Object (Helicobacter pylori) Detection. Am J Clin Pathol 2018;149:181-185. https://doi.org/10.1093/ajcp/aqx154

34 Hood D, Learn DB. Consistent Warthin-Starry Staining Technique. Journal of Histotechnology 1996;19:339-340. https://doi.org/10.1179/1996.19.4.335
35 Stolte M, Meining A. The updated Sydney system: classification and grading of gastritis as the basis of diagnosis and treatment. Can J Gastroenterol 2001;15:591-598. https://doi.org/10.1155/2001/367832

36 Williams C, McCOLL KEL. Review article: proton pump inhibitors and bacterial overgrowth. Alimentary Pharmacology & Therapeutics 2006;23:3-10. https://doi.org/10.1111/j.1365-2036.2006.02707.x

37 Campanella G, Hanna MG, Geneslaw L, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat Med 2019;25:1301-1309. https://doi.org/10.1038/s41591-019-0508-1

38 Iizuka O, Kanavati F, Kato K, et al. Deep Learning Models for Histopathological Classification of Gastric and Colonic Epithelial Tumours. Sci Rep 2020;10:1504. https://doi.org/10.1038/s41598-020-58467-9