Title: Whole-slide imaging: cytomorphological descriptive capacity and intra-observer agreement in canine lymphoma samples

Article Type: Original article

Keywords: Whole slide imaging; Telepathology; Cytology; Lymphoma; Dogs.

Corresponding Author: Dr. Federico Bonsembiante, Ph.D

Corresponding Author's Institution: University of Padua

First Author: Federico Bonsembiante, Ph.D

Order of Authors: Federico Bonsembiante, Ph.D; Valeria Martini; Ugo Bonfanti; Giacomo Casarin; Davide Trez; Maria Elena Gelain

Abstract: The term 'whole-slide imaging (WSI)' refers to the use of robotic microscopes for computerising entire slides into digital images. This study aimed to assess the agreement between WSI and optical microscopy for evaluating canine lymphoma cytological samples. Forty-four slides were computerised using a WSI scanner. The digital and glass slides were examined by three observers with different levels of expertise. The morphology of neoplastic cells and the lymphoma grade were scored, on the basis of the updated Kiel classification, and the intra-observer agreement was assessed. Moreover, the accuracy of determining the lymphoma grade of the digital and glass slides based on the results of flow-cytometry (FC) was established.

The overall intra-observer agreement for the cytomorphological features was fair to moderate (from 0.34 to 0.52) for the three observers, whereas the intra-observer agreement for the evaluation of malignancy grade was moderate (from 0.44 to 0.53). The diagnostic agreement between FC and digital slides was slight (0.16) for the inexperienced observer, fair (0.32) for the mildly experienced observer, and moderate (0.50) for the experienced observer. The diagnostic agreement between FC and glass slides was fair (0.37) for the inexperienced observer, substantial (0.63) for the mildly experienced observer, and moderate (0.50) for the experienced observer. Our findings underline the importance of observer experience in determining the malignancy grade, especially if digital slides are used. The study also identifies some technical limitations of the tested WSI scanner, mainly linked to image quality, which affected the morphological evaluation of the neoplastic cells.
Whole-slide imaging: cytomorphological descriptive capacity and intra-observer agreement in canine lymphoma samples

F. Bonsembiante a,*, V. Martini b, Ugo Bonfanti c, G. Casarin a, D. Trez a, M. E. Gelain a

a Department of Comparative Biomedicine and Food Science, University of Padua, Viale dell’Università 16, 35020 Legnaro (PD), Italy.

b Department of Veterinary Medicine, University of Milan, Via Celoria 10, 20133 Milan, Italy.

c La Vallonèa Veterinary Diagnostic Laboratory, Via Sirtori 9, 20017 Passirana di Rho (MI), Italy.

* Corresponding author. Tel: +39 049 8272518
E-mail address: federico.bonsembiante@unipd.it (F. Bonsembiante)
Abstract

The term ‘whole-slide imaging (WSI)’ refers to the use of robotic microscopes for computerising entire slides into digital images. This study aimed to assess the agreement between WSI and optical microscopy for evaluating canine lymphoma cytological samples. Forty-four slides were computerised using a WSI scanner. The digital and glass slides were examined by three observers with different levels of expertise. The morphology of neoplastic cells and the lymphoma grade were scored, on the basis of the updated Kiel classification, and the intra-observer agreement was assessed. Moreover, the accuracy of determining the lymphoma grade of the digital and glass slides based on the results of flow-cytometry (FC) was established.

The overall intra-observer agreement for the cytomorphological features was fair to moderate (from 0.24 to 0.52) for the three observers, whereas the intra-observer agreement for the evaluation of malignancy grade was moderate (from 0.44 to 0.53). The diagnostic agreement between FC and digital slides was slight (0.16) for the inexperienced observer, fair (0.32) for the mildly experienced observer, and moderate (0.50) for the experienced observer, whereas the diagnostic agreement between FC and glass slides was fair (0.37) for the inexperienced observer, substantial (0.63) for the mildly experienced observer, and moderate (0.50) for the experienced observer. Our findings underline the importance of observer experience in determining the malignancy grade, especially if digital slides are used. The study also identifies some technical limitations of the tested WSI scanner, mainly linked to image quality, which affected the morphological evaluation of the neoplastic cells.

Keywords: Whole-slide imaging; Telepathology; Cytology; Lymphoma; Dogs.
**Introduction**

Digital pathology is a branch of pathology in which images are visualised on a computer monitor rather than directly through a microscope (Weinstein, 1986). Telepathology has evolved from the transmission of static images captured using microscope-mounted cameras to the use of robotic microscopes controlled by pathologists at distant sites, and, more recently, to whole-slide imaging (WSI) (Webster and Dunstan, 2014). A WSI scanner is a robotic microscope capable of digitising an entire glass slide (GS) by using a software to merge the individually captured images into a composite digital image (Pantanowitz et al., 2013). The digital images can be viewed either on the computer used to scan the slides, by using a specific software designed to emulate a light microscope (Webster and Dunstan, 2014), or at a remote site via a high-speed internet connection (Steinberg and Ali, 2001). WSI maintains the relative simplicity of static-image transfer and eliminates its limitations by making available the entire specimen for review (Wilbur, 2011) and allowing the magnification of the digital slide (DS) to be changed (Wilbur et al., 2009). Moreover, the most updated scanners have the capability to perform multiple line scans of the same area at different fields of focus (Webster and Dunstan, 2014). This function, the so-called ‘z-stack’ mode, is most important in acquiring images of cytological specimens, in which cells are often arranged in multiple layers (El-Gabry et al., 2014).

In human medicine, WSI is used for several reasons, mainly digital diagnostics and teleconsultation (Al-Janabi et al., 2012). In veterinary medicine, despite the increased use of WSI instruments in reference laboratories, few reports on digital pathology are present in the literature. To the authors’ knowledge, no papers focusing on the morphological descriptive capability of WSI have been published. Moreover, the use of WSI has never been validated and its capability to replace the conventional microscope has not been assessed using cytological samples.
One of the most commonly obtained cytological samples in canine medicine is the fine needle aspirate (FNA) of lymph nodes. This technique is used to discriminate between inflammatory and neoplastic disorders, among which, non-Hodgkin’s lymphoma is the most common primary neoplasia affecting the lymph node (Richards and Suter, 2015). The identification of a clonal expansion of neoplastic lymphocytes is essential for lymphoma diagnosis. Histology, immunocytochemistry, immunohistochemistry, flow cytometry (FC) and PCR to detect clonal antigen receptor gene rearrangement (PARR) are fundamental tools that can be used to confirm the diagnosis of lymphoma and to identify the phenotype for prognostic purposes (Burkhard and Bienzle, 2013). These diagnostic procedures can be performed only in specialized laboratories.

Commonly, the cytological examination of lymph node samples obtained via FNA is the first step in the diagnosis of lymphoproliferative diseases in dogs, because it is minimally invasive, inexpensive, and fast (Amores-Fuster et al., 2015). Moreover, the cytological features are used to classify the different lymphoma subtypes with prognostic significance (Ponce et al., 2004).

However, the morphological evaluation of neoplastic lymphocytes is complex, and the opinion of a skilled clinical pathologist is often required for the correct diagnosis and classification of the different lymphoma subtypes.

The possibility to use WSI instruments could improve the quality of cytological services and could provide the possibility to share cytological samples of canine lymphomas with more experienced clinical pathologists. However, before this technology can be applied in veterinary cytology, the reliability of cytological evaluation and the technical aspects of the scanning process must be evaluated. The aims of this study were to assess the following: 1) the intra-observer agreement (IOA) between WSI (D-sight, A. Menarini Diagnostics S.r.l) and optical microscopy in the evaluation of cellular morphology in canine lymphoma samples, by using the updated Kiel classification (Ponce et al. 2010); 2) the IOA between WSI and optical microscopy in the assessment of lymphoma grading; and 3) whether the accuracy of grading assessment varied...
between WSI scanner and optical microscopy, by using FC as a reference method. Moreover, the influence of the level of observer expertise on their performance was evaluated.

Materials and methods

Cytological samples and observers

The database of the Flow Cytometry Service of the Department of Veterinary Medicine (University of Milan, Milan, Italy) was searched to select consecutive canine lymphoma samples diagnosed on the basis of clinical, clinico-pathological, cytological, and FC data, from January 2015 to June 2015. Only cases with good-quality lymph node cytological smears were enrolled in the study to allow for a detailed evaluation of the morphological features of the neoplastic cells. FC analysis and the criteria for lymphoma diagnosis were applied as previously described (Gelain et al., 2008). The following antibodies were used: CD45 (clone YKIX716.13, Serotec), CD3 (clone CA17.2A12, Serotec), CD5 (clone YKIX322.3, Serotec), CD4 (clone YKIX302.9, Serotec), CD8 (clone YCATE55.9, Serotec), CD21 (clone CA21D6, Serotec), CD79a (clone MCA1298F, Serotec), and CD34 (clone 1H6, Pharmingen, BD Biosciences). Neoplastic cells were identified on the morphological cytogram (forward scatter [FSC] versus side scatter) or on the CD45 versus FSC cytogram. The percentage of neoplastic cells, the phenotype, and the mean FSC (mean cellular size) were used to identify the cell type (B-cell or T-cell) and cell size. GS were stained with May-Grünwald-Giemsa stain, and for each case, the slide with higher cellularity and better preservation was selected.

Three observers with different levels of expertise participated in the study, and all were blinded to the FC results. The inexperienced observer, a PhD student, had the lower experience in cytological evaluation; the mildly experienced observer, a postdoctoral researcher, had intermediate experience in cytological evaluation; and the experienced observer, a board-certified clinical-
pathologist, had extensive experience in cytological evaluation. None of the observers had previously used the WSI technology for routine diagnostic procedures.

Digital slides

All the GSs were scanned using the 40× objective with the z-stack modality by using a WSI scanner (D-sight, A. Menarini Diagnostics S.r.l.) to obtain the DSs. The DSs were scanned using automated tissue detection and focus-point assignments with seven-line scans of the same area at different fields of focus. The research case numbers assigned to the DSs were different from those assigned to the corresponding GSs to minimize recall bias. DSs were subsequently uploaded to a server to be evaluated by the three observers using an online software (Telepathology, Visia Imaging S.r.l.) (Fig. 1). The monitors used ranged from 14 to 15.6 inches, with a screen resolution of at least 1366×768 pixels, and no special monitor or setting was used. Each observer used his own laptop’s screen to evaluate all the DSs. Between the evaluation of GS and DS by the same observer, a wash-out period of at least one month was imposed to ensure the observers did not remember the cases from the previous viewings.

Morphological features evaluated

According to the updated Ki-67 classification (Fournel-Fleury et al., 1997; Ponce et al., 2010) the following parameters were recorded for both the DSs and GSs by each observer at 40× magnification: pleomorphism of the neoplastic population, cellular size, amount and colour of cytoplasm, nuclear shape, nuclear chromatin pattern, and nucleoli. Specifically, the following features were evaluated:

a. The cellular size was based on the comparison between the red blood cells and nucleus of the cells, and was defined as small, medium, or large depending on whether it was smaller, equal to, or larger than two erythrocytes, respectively.
b. Pleomorphism of the neoplastic population was recorded if the cellular size and shape varied among the neoplastic population.

c. Cytoplasm amount, distribution and colour: the cytoplasm was described as scarce—when only a scant amount of cytoplasm was present—moderate or abundant. The presence of unipolar distribution was also recorded. The colour of the cytoplasm was identified as clear, basophilic or deeply basophilic.

d. Nuclear shape and nuclear chromatin pattern: round, indented or convoluted and irregular shape of the nucleus was recorded, and the chromatin was described as dense, granular or finely granular, or smooth.

e. Presence of nucleoli: the presence of single or multiple prominent nucleoli was recorded.

Before the beginning of the study, the observers conferred to standardize the appearance of all these morphological features.

Moreover the observers were asked to identify the malignancy grade (high or low grade). Small-cell lymphomas were classified as low-grade lymphomas while large-cell and pleomorphic mixed-small-large-cell lymphomas were classified as high-grade lymphomas.

All these morphological features were scored as shown in Table 1.

Statistical methods

In all the cases, the IOA between the DSs and GSs was assessed for each of the morphological features and the grading. Moreover, the agreement between FC and DS and between FC and GS in determining the malignancy grade was also evaluated for each observer to evaluate the accuracy of the DS and GS, respectively, by using FC as a reference method for determining of the lymphoma grade. The IOA and the agreement in determining the malignancy grade were assessed by using linearly weighted Cohen’s K. The K coefficients were interpreted as recommended by Landis and Koch (1977): <0.00, poor; 0.00-0.20, slight; 0.21-0.40, fair; 0.41-0.60,
Differences in the number of correctly classified lymphomas obtained by evaluating the DSs and GSs, compared to the results of FC, were assessed using Fisher’s exact test. Statistical analysis was performed using a commercially available software program (MedCalc Statistical Software version 15.8, MedCalc Software bvba).

Results

In total 250 samples arrived to the Flow Cytometry Service of the Department of Veterinary Medicine, University of Milan from January to June 2015 and 163 of these (65%) also had corresponding cytological slides. In 44 out of the 163 (27%) samples, the cytological smears had good cellularity and preservation and were hence included in the present study. Based on the phenotype and size of cells assessed using FC, 24 high-grade B-cell lymphomas, four low-grade B-cell lymphomas, 10 high-grade T-cell lymphomas, and six low-grade T-cell lymphomas were included in the study (Table 2).

The IOA results for the cytomorphological features are listed in Table 3. For the inexperienced observer the agreement between the DSs and GSs was slight for the amount and colour of cytoplasm, and fair for all the other morphological features. The mean IOA for cytomorphological features was fair. For the mildly experienced observer, the agreement between the DSs and GSs was slight for cellular size, chromatin pattern, nuclear shape, and cellular pleomorphism; fair for the amount of cytoplasm; and moderate for cytoplasm colour and the presence of nucleoli. The mean IOA for cytomorphological features was fair. For the experienced observer the agreement between the DSs and GSs was fair for chromatin pattern, cytoplasm amount, and nuclear shape; and moderate for cellular pleomorphism, cellular size, cytoplasm colour, and the presence of nucleoli. The mean IOA for cytomorphological features was moderate.

For each cytomorphological feature, the mean IOA of the three observers was fair, except for
cytoplasm amount for which it was slight, and for cytoplasm colour and the presence of nucleoli for which it was moderate.

Even if no statistical differences were present, the use of WSI brought about an improvement in the percentage of high-grade lymphomas correctly identified by the mildly experienced observer and in the percentage of low-grade lymphomas correctly identified by the inexperienced observer. Nevertheless, it affected the percentage of high-grade lymphomas identified by the inexperienced and experienced observers and the percentage of low-grade lymphomas identified by the mildly experienced observer (Table 2). The agreement between the FC and DS results in the assessment of grading was slight for the inexperienced observer, fair for the mildly experienced observer, and moderate for the experienced observer, while the agreement between the FC and GS results in the assessment of grading was fair for the inexperienced observer, substantial for the mildly experienced observer, and moderate for the experienced observer (Table 4).

Discussion

In this study, we assessed the reliability of the cytomorphological evaluation of canine lymphoma samples by using a WSI scanner, and found a low IOA between the DSs and GSs for all three observers for the morphological features assessed. The agreement between the DSs and FC was slight for the inexperienced observer and fair to moderate for the two observers with more experience. This underlines the importance of observer experience in the cytological evaluation of lymphoma samples.

In recent years, researchers have shown increasing interest in both human and veterinary digital pathology (Maiolino et al., 2006; Kelly, 2007; Al-Janabi et al., 2012; Webster and Dunstan, 2014; Bertram et al., 2018). The three-dimensional architecture of cells in FNA samples limited the
application of digital pathology to cytological samples. However, with the introduction of WSI scanners with the z-stack function, FNA samples have become more suitable for digitalisation and visualisation on a monitor (El-Gabry et al., 2014). Nevertheless, no studies to date have validated the use of WSI scanners in veterinary cytology. In our study, we chose cytological samples of canine lymphoma to test our WSI scanner for many different reasons: first, lymphoma is the most common haematopoietic tumour in dogs (Richards and Suter, 2015); second, the first-step in diagnosis is often a FNA cytology of the lymph nodes; third, neoplastic lymphoid cells have peculiar morphological features that can be appreciated by cytological analysis (Fournel-Fleury et al., 2002; Ponce et al., 2003). To assess the concordance of information provided by the DSs and GSs we evaluated the intra-observer variability, which is the preferred measure of performance in digital pathology (Thrall et al., 2015). Moreover, to evaluate the diagnostic performance of our WSI scanner, we compared the results of FC analysis to the data provided by the DS and GS analyses for assessing the lymphoma grade on the basis of the fact that, nowadays, FC is considered a fundamental tool in lymphoma diagnosis (Comazzi and Gelain, 2011; Comazzi et al., 2016). The aim of this comparison was to evaluate whether the possible disagreement was due to the descriptive capacity of the DSs or whether the challenge in the morphological identification of grading was intrinsic to the cytological examination.

In the evaluation of morphological features, the overall IOA of the three observers was fair to moderate. These IOAs are lower than those reported in the validation studies of other WSI scanners used for histopathological analysis in veterinary pathology (Bertam et al., 2018) and for histopathological (Thrall et al., 2015), cytological (House et al., 2013), and haematological (Gomez-Gelvez et al., 2015) analyses in human pathology. The lower agreement obtained in our study is likely not only because of the shortcomings of the tested WSI scanner, but also because of the different design of our study compared to that of other studies in the literature. Indeed, the main goal of our study was to determine the descriptive capacity of the WSI scanner rather than its
‘diagnostic performance’. The evaluation of morphological features of neoplastic lymphocytes could also be difficult when using an optical microscope, and an unequivocal concordance on single morphological features is challenging. However, our results clearly indicate that observer experience is important in the cytological evaluation of lymphomas, as shown by the increasing IOA from the inexperienced observer to the experienced one. Nevertheless, some technical limitations of our WSI scanner contributed to the low level of agreement. In particular, the lowest mean IOA was recorded in the evaluation of cytoplasm amount. The online software interface, despite the z-stack scanning modality, does not have the capability to ‘focus’ up and down and does not allow increasing the brightness of the DSs, thereby leading to a darker background than that obtained for the GSs. The lack of these functions caused difficulties in the assessment of the edges of the cells. Moreover, the cytoplasm was considered more frequently basophilic or deeply basophilic in the DS than in the GSs (Fig. 2). Conversely, the evaluation of the nucleoli was not conditioned by the lack of these functions, and the presence of nucleoli was the microscopic detail with the highest IOA. Thus, these limitations did not completely hamper our ability to evaluate nuclear morphology.

The mitotic count (MC) is defined as the number of mitoses in 10 consecutive high-power fields (Meuten et al., 2016), and it is a useful parameter to define the malignancy grade of lymphomas (Ponce et al., 2010). MC in cytology is somewhat controversial because of the uneven distribution of cells throughout a sample, but it is considered a more reproducible and reliable tool in histopathological examinations (Sapierzyński et al., 2016). Moreover, in our study, even if the magnification used to analyse the DSs and GSs was the same, the area occupied by the cells in the DSs viewed on the monitor was smaller than that of cells in the GSs, because a portion of the monitor is occupied by the online software interface with navigation icons and commands. Therefore, the number of cells/40× field of a DS was lower than that of cells/40× field of a GS, and thus, the different number of cells/40× field did not allow for a comparison of the MC between the
DSs and GSs. An alternative approach to comparing mitosis between the two methods is to consider the mitotic index (MI), which is the ratio of the number of cells in mitosis and the number of cells not in mitosis (Meuten et al., 2016). Nevertheless, this approach is time-consuming because it is necessary to count the cells not in mitosis, usually 1,000, and it is not applicable in the routine diagnostic practice. For all these reasons, the MI was not included as a malignancy feature in our study.

In addition to the possibilities to diagnose a lymphoma, the cytological examination of lymph node FNAs could be useful to predict the grading of the neoplastic process. Large cells and/or cellular pleomorphism could be suggestive of high grade lymphomas (e.g. centroblastic B-cell lymphoma or pleomorphic large T-cell lymphoma), while small cells are more frequently observed in low grade lymphomas (e.g. small lymphocytic B-cell lymphoma or small clear T-cell lymphoma) (Fournel-Fleury et al., 2002; Ponce et al., 2010). Despite the technical limitations of our WSI scanner that led to difficulties in the evaluation of some morphological features, the IOA in the determinations of grading was moderate for all the observers. These results may suggest that despite some difficulties in the evaluation of some specific characteristics, it is possible to obtain acceptable results by using digital smears even with non-optimal images.

The first aim of FC exam is the objective determination of the lymphoma’s phenotype. However, the information derived from FC go far beyond the simply identification of cell origin: it allows the assessment of the percentage of neoplastic cells within FNA sample, their size, based on the mean FSC, and the pleomorphism, based on the standard deviation of the FSC (Gelain et al., 2008). Moreover, it is possible to evaluate the phenotype aberrations (both qualitative and quantitative) which, in combination with morphological features (e.g size) allows to recognized some specific lymphoma subtypes (e.g small clear cell lymphoma). In our study, the lymphoma’s grade determined by flow cytometry was compared to the lymphoma’s grade obtained with DS and...
GS by the three observers. Regarding lymphoma grade, the results of the agreement with FC have shown a correlation with observer experience. In fact, agreement was slight and fair for the DSs and GSs, respectively, for the inexperienced observer, thus reflecting the observer’s difficulty in determining the lymphoma grade based only on the morphological features of the cells. In contrast, the mildly experienced observer had good agreement when using the optical microscope and a slightly decreased accuracy when using the WSI scanner, and the experienced observer had moderate agreement when using both the DSs and GSs. These data confirm the importance of observer experience in the evaluation of lymphoma samples and reflect the lower descriptive capacity of the DSs than the GSs.

From a technical perspective, the scanning time for each DS was between three and four hours, depending on the amount of material on the GS. The time interval is related to the z-stack scanning mode with 40× magnification. The use of WSI scanners with such long scanning times negatively affects their application in routine cytological diagnosis, wherein a single case often has multiple GSs. However, when challenging cases require the opinion of skilled clinical-pathologists at distant sites, this technology can be an efficient tool. To simulate a routine workflow in the use of the WSI scanner, for each case, we chose the slide with higher cellularity and better preservation, because high cellularity and good preservation are characteristics that a regular practitioner can also recognize.

The limitation of the long scanning time could be solved using a more advanced WSI scanner with a higher speed of acquisition. The size of the DS (from 1 to 2 gigabytes) could also be a limitation and the use of a dedicated server is mandatory to facilitate the storage and sharing of the files (Gomez-Gelvez et al., 2015).

Conclusions
The data from our study underline some technical limitations of the tested WSI scanner, mainly linked to image quality, which could limit the diagnostic power of the instrument. We have also demonstrated the importance of observer experience in the correct interpretation of DSs. However, given the technological advances and development of new functions, the digital cytological workflow in veterinary medicine could also be improved.

**Conflict of interest statement:** None of the authors of this paper have a financial or personal relationship with other people or organisations that could inappropriately influence or bias the content of the paper.

**Acknowledgements**

The WSI scanner (D-sight, A. Menarini Diagnostics S.r.l.) was acquired thanks to the scientific instrumentation grant (2012) by the University of Padua.

The authors gratefully acknowledge the assistance of Dr. Alessandra Faggionato and of Dr. Francesco Cian in English editing.

The preliminary results of this study were presented as an Abstract at the 18th Congress of the European Society of Veterinary Clinical Pathology, Nantes, 20-22 October 2016.

**References**


Morphological features and grading system used in the cytological evaluation.

<table>
<thead>
<tr>
<th>Morphological features</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cellular size</td>
<td>Small</td>
<td>Medium</td>
<td>Large</td>
<td></td>
</tr>
<tr>
<td>Cellular pleomorphism</td>
<td>Present</td>
<td>Absent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cytoplasm amount</td>
<td>Scarce</td>
<td>Moderate</td>
<td>Abundant</td>
<td>Unipolar</td>
</tr>
<tr>
<td>Cytoplasm colour</td>
<td>Clear</td>
<td>Basophilic</td>
<td>Deeply basophilic</td>
<td>Irregular</td>
</tr>
<tr>
<td>Nuclear shape</td>
<td>Round</td>
<td>Indented</td>
<td>Convoluted</td>
<td></td>
</tr>
<tr>
<td>Chromatin pattern</td>
<td>Dense/Thickened</td>
<td>Granular/Coarse</td>
<td>Finely granular</td>
<td>Smooth</td>
</tr>
<tr>
<td>Nucleoli</td>
<td>Not present</td>
<td>Single</td>
<td>Multiple</td>
<td></td>
</tr>
</tbody>
</table>
Table 2
Classification of the 44 lymphoma samples based on the flow cytometry results and the results of evaluation of the digital and glass slides by the three observers.

<table>
<thead>
<tr>
<th></th>
<th>High grade lymphomas</th>
<th>Low grade lymphomas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flow cytometry</td>
<td>34 (24 B-cells; 10 T-cells)</td>
<td>10 (4 B-cells; 6 T-cells)</td>
</tr>
<tr>
<td>Observer 1 - digital slides</td>
<td>23</td>
<td>9</td>
</tr>
<tr>
<td>Observer 2 - digital slides</td>
<td>30</td>
<td>5</td>
</tr>
<tr>
<td>Observer 3 – digital slides</td>
<td>27</td>
<td>7</td>
</tr>
<tr>
<td>Observer 1 – glass slides</td>
<td>26</td>
<td>8</td>
</tr>
<tr>
<td>Observer 2 – glass slides</td>
<td>28</td>
<td>8</td>
</tr>
<tr>
<td>Observer 3 – glass slides</td>
<td>30</td>
<td>7</td>
</tr>
</tbody>
</table>

Comment [FB290]: Line 435 of YTVJL-D17-00838R2 Deleted – using digital and glass slides

Comment [FB291]: Table 2 of YTVJL-D17-00838R2 Deleted - High grade B-cell lymphomas Low grade B-cell lymphomas High grade T-cell lymphomas Low grade T-cell lymphomas Flow cytometry 24 4 10 6 Observer 1 - digital slides 12 4 2 4 Observer 2 - digital slides 18 1 7 3 Observer 3 – digital slides 21 1 2 5 Observer 1 – glass slides 16 2 1 4 Observer 2 – glass slides 18 1 8 6 Observer 3 – glass slides 23 1 7 6
Intra-observer agreement for morphological features, grading, and phenotype. Observer one corresponded to the lower level of cytological experience; observer two corresponded to the intermediate level of cytological experience; observer three corresponded to the higher level of cytological expertise. The intra-observer agreement was assessed by using linearly weighted Cohen’s K. The K coefficients were interpreted as recommended by Landis and Koch (1977).

<table>
<thead>
<tr>
<th>Morphological features</th>
<th>Coefficient K observer 1 (95% CI)</th>
<th>Coefficient K observer 2 (95% CI)</th>
<th>Coefficient K observer 3 (95% CI)</th>
<th>Mean coefficient K (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cellular size</td>
<td>0.29 (-0.07 - 0.50)</td>
<td>0.00 (-0.26 - 0.21)</td>
<td>0.44 (0.14 - 0.72)</td>
<td>0.30 (0.16 - 0.44)</td>
</tr>
<tr>
<td>Cellular pleomorphism</td>
<td>0.24 (-0.04 - 0.52)</td>
<td>0.17 (-0.10 - 0.43)</td>
<td>0.41 (0.14 - 0.68)</td>
<td>0.29 (0.13 - 0.45)</td>
</tr>
<tr>
<td>Cytoplasm amount</td>
<td>0.03 (-0.16 - 0.21)</td>
<td>0.26 (0.04 - 0.47)</td>
<td>0.28 (0.04 - 0.51)</td>
<td>0.17 (0.04 - 0.31)</td>
</tr>
<tr>
<td>Cytoplasm colour</td>
<td>0.13 (-0.10 - 0.36)</td>
<td>0.55 (0.35 - 0.75)</td>
<td>0.43 (0.21 - 0.64)</td>
<td>0.43 (0.30 - 0.56)</td>
</tr>
<tr>
<td>Nuclear shape</td>
<td>0.28 (0.08 - 0.48)</td>
<td>0.14 (-0.23 - 0.52)</td>
<td>0.33 (-0.03 - 0.69)</td>
<td>0.31 (0.16 - 0.47)</td>
</tr>
<tr>
<td>Chromatin pattern</td>
<td>0.32 (0.10 - 0.55)</td>
<td>0.03 (-0.13 - 0.19)</td>
<td>0.22 (0.03 - 0.40)</td>
<td>0.22 (0.11 - 0.34)</td>
</tr>
<tr>
<td>Nucleoli</td>
<td>0.33 (0.11 - 0.55)</td>
<td>0.52 (0.21 - 0.72)</td>
<td>0.50 (0.25 - 0.75)</td>
<td>0.48 (0.36 - 0.60)</td>
</tr>
<tr>
<td>cytomorphic features</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.34 (0.26 - 0.42)</td>
<td>0.40 (0.32 - 0.48)</td>
<td>0.52 (0.44 - 0.60)</td>
<td></td>
</tr>
<tr>
<td>Malignancy grade</td>
<td>0.53 (0.29 - 0.78)</td>
<td>0.44 (0.10 - 0.78)</td>
<td>0.46 (0.18 - 0.74)</td>
<td>0.51 (0.34 - 0.70)</td>
</tr>
</tbody>
</table>
Agreement between flow-cytometry results and the digital and glass slide assessments of lymphoma grading. Observer one corresponded to the lower level of cytological experience; observer two corresponded to the intermediate level of cytological experience; observer three corresponded to the higher level of cytological expertise. The diagnostic agreement was assessed by using linearly weighted Cohen’s K. The K coefficients were interpreted as recommended by Landis and Koch (1977).

<table>
<thead>
<tr>
<th></th>
<th>Coefficient K observer 1 (95% CI)</th>
<th>Coefficient K observer 2 (95% CI)</th>
<th>Coefficient K observer 3 (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flow cytometry results vs digital slide assessments</td>
<td>0.16 (-0.13-0.50)</td>
<td>0.32 (0.06-0.58)</td>
<td>0.50 (0.23-0.77)</td>
</tr>
<tr>
<td>Flow cytometry results vs optical slide assessments</td>
<td>0.37 (0.12-0.62)</td>
<td>0.63 (0.39-0.87)</td>
<td>0.50 (0.24-0.75)</td>
</tr>
</tbody>
</table>
Figure legends

Fig. 1. Whole-slide imaging microscope imaging workstation (D-sight, A. Menarini Diagnostics S.r.l.). 1a. The system consists of a five-slide capacity scanner with four objectives (4×, 10×, 20×, and 40×) and a high-performance desktop computer. 1b. The main page of the online software used for analysis (Telepathology, Visia Imaging S.r.l). The system allows to digitise, store, and preview all digitalised slides. Thus, the user can scroll through the archive and select the slide of interest. 1c. The online software’s navigation page: using the image multi-preview, the user can select and edit any area of interest (black asterisk). Information regarding the different magnifications available (white asterisk), the current magnification (yellow asterisk), and the navigation map of the area selected (red asterisk) are also present.

Fig. 2. Small-clear cell lymphoma: 2a. Digital slide and 2b. Glass slide of the same sample (40× magnification). Compared to the glass slide, the digital slide, clearly shows the darker background and, consequently, the more basophilic cytoplasm of the cells.
Highlights

- Whole slide imaging scanner is a robotic microscope capable to digitize a glass slide to create a composite digital image
- The cytomorphological capability of a whole slide imaging scanner is evaluated on 44 lymphoma samples
- There are some technical limitations that affected the diagnostic power of the instrument
- The release of new functions will be important to overcome the technical limitations
- This is the first time in which a whole slide imaging scanner is evaluated in Veterinary Medicine
Revision Note
Manuscript Number: YTVJL-D-17-00838R2

Article Title: Whole-slide imaging: cytomorphological descriptive capacity and intra-observer agreement in canine lymphoma samples

REVIEWERS' COMMENTS:

Reviewer #1:
---------------------------------

**Reviewer comment** - Thanks for addressing most of the comments and concerns that were raised in the review process. I did note that the descriptive breakdown of agreement for morphological criteria were removed from the results in this version. I understand that the kappa values are listed in the tables, but I appreciated the information in the text of the previous version as well.

**Author response** – We added the requested information in the main manuscript (lines 182-194).

**Reviewer comment** - Regarding the impact of slide quality on whole slide imaging, I maintain that it would be valuable to have at least 1 sentence stating the percentage of samples that satisfactorily passed the QC process. According to the numbers provided in the rebuttal, only 27% of samples passed were of high enough quality for scanning. This is an important consideration for those that are interested in this technology.

**Author response** – We added this information in the main manuscript (lines 174-177).

Minor comments:

**Reviewer comment** - Page 10, Line 242: I’m sorry I didn’t pick up on this with the first review, but is “server” the correct word here? Is the focusing an issue with the server or the interface?

**Author response** – We changed ‘server’ with ‘online software interface’ (line 251).

Reviewer comment - Page 12, Line 281: Wasn’t the observer 2’s IOA for phenotype substantial, not moderate?

**Author response** – We deleted all the information regarding the phenotype from the text, according to Reviewer 2 and Editor’s comments.

Reviewer #2:
---------------------------------

The authors have addressed many of my concerns and the paper is much improved. A few minor issues should still be addressed, in addition to some general editing for language to improve understanding.

**Reviewer comment** - Lines 73-75: Not all those tests are required to confirm lymphoma, therefore I would not say they are all fundamental. Consider rephrasing to “are fundamental tools that can be used to confirm and type lymphoma, etc.” or something like that.

**Author response** – We added ‘tool that can be used’ in the main text (line 73).

**Reviewer comment** - Most instances of the word “telepathology” should be changed to “digital pathology” as no one was remotely controlling a microscope.

**Author response** – We changed ‘telepathology’ with ‘digital pathology’ in the main text (lines 43, 62, 217, 219, 229).

**Reviewer comment** - I remain in opposition to presumptively phenotyping lymphoma based on cytological morphology alone. I am concerned that this would give the impression that cytologic morphology is sufficient to type lymphoma when it is not. The publications used to justify classifying
immunophenotype based on cytological criteria did not actually use cytology as the gold standard for tumor classification. Ponce et al 2010 only examined cytology on 93/608 cases and did not use the cytology to classify the tumor types; the “cytological criteria” in their tables are the criteria used for examining histopathology slides not cytology slides. Similarly with Ponce 2004, the immunophenotype was determined by immunohistochemistry and immunocytochemistry, i.e. not based on cytological morphology. If this is kept, I would prefer to see a stronger emphasis in the discussion (and abstract) that cytological classification of lymphoma type is presumptive/preliminary and definitive methods (e.g. flow, IHC) are REQUIRED for true classification.

**Author response** – We agree with the reviewer about the needs to use immunological method to confirm the phenotype of lymphomas and we underlined it both in the text and in the revision note in the previous revision. We define the morphological evaluation only “an attempt “to determine the phenotype based on what is reported in literature (see Fournel-Fleury et al., 2002), being absolutely aware of the difficult to correctly correlate morphological features and immunophenotype in these neoplasms. However, according to the revisions of the Reviewer 2 and of the Editor, we deleted the part regarding the determination of the phenotype using cytology from the main text.

**Reviewer comment** - In my opinion, comparing morphological criteria identified on DS and GS vs FC to classify lymphomas is not so much evaluation of the ability of cytologists to identify lymphoma types per se, as it is an evaluation of the correlation between the evaluated morphological features and lymphoma immunophenotypes. Please comment.

**Author response** – Undoubtedly, the first aim of the flow cytometry exam is the objective determination of the lymphoma's phenotype. However, the information derived from FC go far beyond the simply identification of cell origin: it allows the assessment of the percentage of neoplastic cells within FNA sample, their size, based on the mean FSC, and the pleomorphism, based on the standard deviation of the FSC; it’s possible to evaluate the phenotype aberration (both qualitative and quantitative) which, in combination with morphological features (e.g size) allows to recognized some specific lymphoma subtypes (e.g small clear cell lymphoma). Taken together, all these data allow to make diagnosis and to identify some lymphoma subtypes. In our study, the lymphoma's grade determined by flow cytometry was compared to the lymphoma’s grade obtained with DS and GS by the three observers. We added this information in the main text (lines 289-297).

**Reviewer comment** - Throughout the paper, I would suggest replacing observer number (1, 2, or 3) with experience level, e.g. “inexperienced”, “mildly experienced” and “experienced” observer. This allows someone reading the paper to understand the context without referring back to the M&M to make that determination.

**Author response** – We replaced observer one with inexperienced observer, observer two with mildly experienced observer, and observer three with experienced observer in the main text.

I think the reasoning in the response to the following earlier reviewer comment is valid and should be included in the discussion:

**Reviewer comment** - Lines 106-107: “most representative slide” automatically creates bias. In a clinical situation, would a regular practitioner know how to choose the best slide? Also, based on reviewer experience, this information may not apply to a regular practitioner who often has less than the equivalent of 2 years cytology expertise.

**Author response** - It is our opinion that the WSI technology could be applied to cytological samples only for second opinion or for selected challenging cases in which the experience of a board-certified clinical-pathologist is required and, due to the long scanning time, not for the routine diagnostic procedures. For these reasons we included only cases with high cellularity and well preserved. The cellularity and preservation of a sample are characteristics that also a regular practitioner can recognize. We changed ‘the most representative slide’ with ‘the slide with higher cellularity and better preservation’ in the revised manuscript (lines 115-116).

**Author response** – We added this information in the discussion section (lines 312-315).
I also think the information below should be added to the M&M:

**Reviewer comment** - Reviewer comment - Lines 130-136: Did cytologists confer before the study to agree on the appearance of the morphological features for standardization within the study?

*Author response* - Yes, the cytologists conferred before the start of the study to agree on the appearance of the morphological features of the cells.

**Author response** – We added this information in the M&M section (line 153-154).

**Reviewer comment** - Lines 143-144: Variation in cell size is anisocytosis, not pleomorphism. Also please define the anisocytosis cut-off at which "pleomorphism" was considered present. Greater than 10% anisocytosis? Greater than 50%? Greater than 100%?

*Author response* – In the cytological evaluation of neoplastic lymph-nodes, the presence of cells with various size and shape is defined 'pleomorphism' (Fournel-Fleury et al., 2002), and the lymphomas are classified as pleomorphic (e.g. centroblastic polymorphic; small-, mixed-, or large-cell pleomorphic non-hodgkin lymphoma) when cells with different size are present (Ponce et al., 2010). In literature no cut-offs at which pleomorphism is considered present are reported.

**Reviewer comment** - Lines 145-148: Was perinuclear clearing considered in the distribution of the cytoplasm colour?

*Author response* – We did not consider the perinuclear halo in the distribution of the cytoplasm colour.

**Reviewer comment** - Line 152: Was variation in size or shape of nucleoli recorded?

*Author response* – We considered the presence/absence and the number of the nucleoli but not the variation in shape and size

**Reviewer comment** - Lines 153-164: What is the value of comparing the classification of low vs high grade lymphoma, when they are the exact same as comparing the classification of cell size (small vs. large and mixed)? Similarly, based on the descriptions provided B vs T cell phenotype is based on potentially dichotomous criteria (clear vs. colored cytoplasm, round vs. not-round nuclei), so why not simply compare the observers' abilities to detect those?

*Author response* – The cellular size is the only objective parameter that was also included in the main text as single morphological feature, but the lymphoma’s grade is determined by an overall assessment of the cytological slides.

**Reviewer comment** - It would be interesting to consider, if the data is available, how the observer’s subjective assessment of cell size (small, medium, large) correlated with the FC measurement of cell size?

*Author response* – Actually, this correlation is not one of the aim of our study, so we didn’t record these data.

**Reviewer comment** - Lines 191-194: Are these changes statistically significant?

*Author response* – The changes are not statistically significant, we added this information in the main text (line 196).

**Reviewer comment** - Line 183/Table 3: What statistical test was used to determine if the differences between observers was statistically significant or not? Alternatively, what are the confidence intervals for these Kappas?

*Author response* – We added the 95% confidence interval for the Kappa values in tables 3 and 4.

**ADDITIONAL EDITORIAL COMMENTS:**
**Reviewer comment** - We agree with Reviewer 2 that (presumptive) phenotyping/immunophenotyping of lymphoma should not be done on cytomorphology alone; this section of the manuscript should be removed.

**Author response** – We deleted the phenotypical assessment based on the morphological evaluation of the neoplastic cells from the main text.

**Reviewer comment** - Please change “Declaration of interest” to "Conflict of interest statement".

**Author response** – Done (line 328).

In addition, one of the reviewers has made the following comments:

**Reviewer comment** - The Title is too long and should be reworded.

**Author response** – The Title has been changed and reduced (lines 3-4).

**Reviewer comment** - The current Abstract is not useful and needs to be rewritten.

**Author response** - We rewrote the abstract (lines 20-39).

**Reviewer comment** - Overall, the standard of English needs to be improved and requires major editing.

**Author response** – The manuscript has been edited by the Elsevier Language Editing Service. The certificate is uploaded together with the other files.

**Reviewer comment** - There is a need to provide an evaluation of the statistical significance of the differences between kappa values.

**Author response** – The K value is a measure of the agreement between two observers using two different techniques or between two techniques used by the same observer. Rarely the statistical significance for Cohen’s K is reported, probably because even relatively low values of K can nonetheless be significantly different from zero but not sufficient magnitude to satisfy researchers (Bakeman and Gottman, 1997). Still, its standard error has been described (Fleiss et al., 1969). We added the 95% CI as requested by Reviewer 2.

**Reviewer comment** - The discussion is repetitive and needs to be streamlined.

**Author response** – We changed the discussion (lines 209-319).
Dear Editor,

Please find enclosed herewith the revised version of the manuscript titled "Whole-slide imaging: cytomorphological descriptive capacity and intra-observer agreement in canine lymphoma samples (YTVJL-D-17-00838R2)".

We answered to all the revisions of the two Reviewers and of the Editor. The title’s length was reduced, the abstract was changed, and all the data regarding the determination of phenotype using cytology exam were removed from the manuscript. The manuscript was revised by the Elsevier’s Language Editing Service (the certificate is attached together with the other files). To add the 95% CI of the K values we changed the statistical program (from SPSS to MedCalc). With the new statistical the K value of some parameters is slightly changed. The manuscript was revised according to the new K values.

Feel free to contact me for any questions.
Thank you in advance.
Best regards.

Federico Bonsembiante, DVM, Ph.D
Dept. Comparative Biomedicine and Food Science
University of Padua
AGRIPOLIS - Viale dell'Università 16
35020 Legnaro (PD) – Italy
Fax : + 39.049.8272973
e-mail : federico.bonsembiante@unipd.it