

A Competitive Massive Slide Scanner Test: *diagnosis performance*

Filippo Crivelli

*Department of Pathology, Azienda Ospedaliera
Ospedale di Circolo di Busto Arsizio (VA)*

Outline

- Rationale
- Study design
- Materials
- Methods
 - operations undertaken
 - analysis
- Results
- Discussion

Outline

This report is based on a preliminary survey on 116 cases (8.89% of total scanned cases) in which a diagnose was made on digital slides.

On a statistical basis, this preliminary report is representative of the whole population under investigation

Rationale

- **Aim:** an evaluation of reproducibility of diagnoses made on digital slides as compared with original diagnoses made by microscope
- **Research questions:**
 - Does a diagnose made on digital slides on a flat screen be different from the same one made on a microscope?
 - 20x magnification as operational standard for scanned slides is reliable enough or can give the wrong diagnose due to the difficult of focusing at higher magnification?
 - May different procedures or methods used to get slides by the different labs originate a bias when slides are scanned and what's the impact on digital diagnoses?
 - Are clinical info regarding the patient and the disease strongly necessary to make a correct second opinion diagnoses or not?

Study design

- From the data base of digital slides, it was randomly selected about 10% of total cases. The images of the slides were cleaned of any reference and were analyzed on an anonymous basis by the pathologists;
- Each case had slides scanned by the three different scanners, totalized a pool of about 1,260 digital case collected in triplets;
- Every case was given to three pathologists team: the team from the same hospital where the original diagnosis was made and to two other teams from different hospitals. At the end of the survey digital diagnoses made were about 3,780, nine for every case;
- The output was evaluated considering the original microscope diagnosis as the gold standard

Materials

- Until now, from the data base of the scanned slides (21,246 equivalent to 4,200 cases), 348 cases/scanner were randomly selected (116 cases from each scanner)
- For the viewing of the selected slides a universal reading system able to deal with all the different images was used (SlidePath);

Materials (cont'd)

- 3 pathologists teams from three different hospital have produced digital diagnoses on the 348 selected cases/scanner;
- 1 senior pathologist has revised each single diagnose with the aim of evaluating its correspondence with the original diagnose made with a microscope.

Methods: operations undertaken

- Packages of digital slides were prepared
- Each hospital has identified some pathologists to carry out the diagnoses on digital slides of the assigned cases
- The cases were assigned randomly to each pathologist
- Each pathologist did a diagnose and filled a CRT with its data and an assessment on the quality of the image
- Correspondence of the digitally based diagnoses with the original ones were evaluated by a senior pathologist

Methods: operations undertaken (cont' d)

- Evaluation of correspondence was made on a score scale from 1 to 10 where 1 was inadequate and 10 full correspondence.
- Subsequently data were grouped in a four level scale as follow:
- 0 = inadequacy (no diagnose possible)
- 1 = wrong diagnose (score 2-3-4)
- 2 = incomplete or inaccurate diagnose (score 5-6-7)
- 3 = correct diagnose (score 8-9-10)

Methods: analysis

- Preliminary data were analyzed in a simple way to evaluate only the diagnoses' correspondence in percentage values
- All the statistical analysis on the complete data are in progress. The data collection will be concluded when all the slides/cases would have been diagnosed by the pathologists of the different hospitals which provided the slides and at least by other pathologists. This will allows to evaluate the real possibility of a digital diagnose, not only as a second opinion.
- Future steps will assign to data grouped by hospital, by pathologist, by case an evaluation and marks on assessment of images and efficiency of each scanner will be collected

Results

| | Scanner A | Scanner B | Scanner C | TOTALS |
|------------------|-----------|-----------|-----------|--------|
| No. of diagnoses | 116 | 116 | 116 | 348 |
| Average level | 2.655 | 2.681 | 2.310 | 2.548 |
| Level 3 | 81% | 83% | 65% | 76% |
| Level 2 | 9% | 8% | 14% | 10% |
| Level 1 | 5% | 4% | 9% | 6% |
| Level 0 | 5% | 5% | 12% | 7% |

Discussion

- the present study provides novel insights into currently available slide scanners on their large scale application in a slide scanning service
- as far as we know, no other study has yet attempted the same kind of intensive evaluation – though some relevant evaluations have been done in the Scanner Contests 2010-2012
- **here we present some preliminary discussion**

Discussion/1: image quality

- Images were generally good; In fact, an average of 76% of fully correct diagnoses were made (82% considering only scanner A and B)
- Even more so, if considered that the type of screens used were common computer flat screens, and not HD wide monitors
- About 7% of images were not good enough to make a diagnose (5% considering only scanner A and B)
 - preparing for scanning is more crucial than preparing for microscope + human eye!
- massive scanning -> **need for preparation standards**

Discussion/2: diagnoses

- About 86% of correct diagnoses were made by different pathologists on a basis of digital images with only 10% of incomplete or inaccurate diagnoses
- Main problems arise from loss of focus in dirty slides, thick sections or overstained sections, no information about antigens in IIC slides.
- Completely wrong diagnoses are only 6%

Discussion/2: diagnoses

- Answers to the research questions
 - Does a diagnose made on digital slides on a flat screen be different from the same one made on a microscope?
 - No substantial differences (up to 90.5% of correct diagnoses)
 - 20x magnification as operational standard for scanned slides is reliable enough or can give the wrong diagnose due to the difficult of focusing at higher magnification?
 - Given the results above the magnification is negligible problem

Discussion/2: diagnoses

- Answers to the research questions
 - May different procedures or methods used to get slides by the different labs originate a bias when slides are scanned and what's the impact on digital diagnoses?
 - Yes on a 12% of slides diagnoses were not possible or difficult due to procedural issues in preparing slides (hence standardization of methods is important)
 - Are clinical info regarding the patient and the disease strongly necessary to make a correct second opinion diagnoses or not?
 - No serious impact on second opinion diagnoses. It cannot be assessed if any impact on original diagnoses

Conclusions/1

- this experimentation of massive scanning in a real world scenario reveals some information useful for everyone interested in applying Digital Pathology to surgical pathology routine
- scanners seem to provide a good image quality and reliable behavior
- no particular difficulties to make a correct diagnoses even without essential informations about patient or previous examinations or essential informations about technical procedures (e.g.: antigens in IIC slides)

Conclusions/2

- the next step further is to complete the diagnoses by different pathologists on the entire data base and to analyze all the data on a scanner, on a pathologist, on a hospital basis and on a case basis to evaluate the real impact of digital scanning on second opinion diagnoses and, more important, on original diagnoses made on digital slides on HD flat screen