Opportunities for Standardization and Collaboration in Developing Histopathology Image Analysis Algorithms

Steven N. Hart, Ph.D
@StevenNHart
DISCLOSURE

In the past 12 months, I have not had any significant financial interest or other relationship with the manufacturers of the products or providers of the services that will be discussed in my presentation.
• Brief review of Digital Pathology Informatics for Whole Slide Images

• Best Practices in Algorithm Development

• Minimum Information about Experiments

• Identify Opportunities for Shared Code/APIs
PROCESS

1. Identify the problem
2. Design an experiment
3. Generate data
4. <<Magic Happens Here>>
5. Publish paper
6. Go to conferences
7. Get Famous
8. ...

1. Read the literature to find out what has already been done
2. Find code that already does most of the work for you
3. Try the simplest way first
4. Clean your data
5. Benchmark models
6. Report findings & code so that others can reproduce your work
Digital Pathology Informatics for Whole Slide Images
STAGES OF WSI ANALYSIS

Identifying datasets
Patch Extraction
Preprocessing & Normalization
Applying A Model
Evaluating the Model
IDENTIFYING DATASETS

- TCIA
- BreakHis
- CAMELYON17
- GTEx
- TUPAC16
- TMAD

* Most are not downloadable
<table>
<thead>
<tr>
<th>Dataset or paper</th>
<th>Image size (px)</th>
<th># images</th>
<th>Stain</th>
<th>Disease</th>
<th>Additional data</th>
<th>Potential usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>KIMIA960 [87,88]</td>
<td>308 × 168</td>
<td>960</td>
<td>H&amp;E/IHC</td>
<td>various tissue</td>
<td></td>
<td>Disease classification</td>
</tr>
<tr>
<td>Bio-segmentation [89,90]</td>
<td>896 × 768, 768 × 512</td>
<td>58</td>
<td>H&amp;E</td>
<td>Breast cancer</td>
<td></td>
<td>Disease classification</td>
</tr>
<tr>
<td>Bioimaging challenge 2015 [91,92]</td>
<td>2040 × 1536</td>
<td>269</td>
<td>H&amp;E</td>
<td>Breast cancer</td>
<td></td>
<td>Disease classification</td>
</tr>
<tr>
<td>Glas [23,93]</td>
<td>574–775 × 430–522</td>
<td>165</td>
<td>H&amp;E</td>
<td>Colorectal cancer</td>
<td>Mask for gland area</td>
<td>Gland segmentation</td>
</tr>
<tr>
<td>BreakHis [15,94]</td>
<td>700 × 460</td>
<td>7909</td>
<td>H&amp;E</td>
<td>Breast cancer</td>
<td></td>
<td>Disease classification</td>
</tr>
<tr>
<td>Jakob Nikolos et al. [88,95]</td>
<td>1000 × 1000</td>
<td>100</td>
<td>IHC</td>
<td>Colorectal cancer</td>
<td>Blood vessel count</td>
<td>Blood vessel detection</td>
</tr>
<tr>
<td>MITOS-ATYPIA-14 [96]</td>
<td>1539 × 1376, 1663 × 1485</td>
<td>4240</td>
<td>H&amp;E</td>
<td>Breast cancer</td>
<td>Coordinates of mitosis with a confidence degree/six criteria to evaluate nuclear atypia</td>
<td>Mitosis detection, nuclear atypia classification</td>
</tr>
<tr>
<td>Kumar et al. [97,98]</td>
<td>1000 × 1000</td>
<td>30</td>
<td>H&amp;E</td>
<td>Various cancer</td>
<td>Coordinates of annotated nuclear boundaries</td>
<td>Nuclear segmentation</td>
</tr>
<tr>
<td>MITOS 2012 [20,99]</td>
<td>2084 × 2084, 2252 × 2250</td>
<td>100</td>
<td>H&amp;E</td>
<td>Breast cancer</td>
<td>Coordinates of mitosis</td>
<td>Mitosis detection</td>
</tr>
<tr>
<td>Janowczyk et al. [100,101]</td>
<td>1388 × 1040</td>
<td>374</td>
<td>H&amp;E</td>
<td>Lymphoma</td>
<td>None</td>
<td>Disease classification</td>
</tr>
<tr>
<td>Janowczyk et al. [100,101]</td>
<td>2000 × 2000</td>
<td>311</td>
<td>H&amp;E</td>
<td>Breast cancer</td>
<td>Coordinates of mitosis</td>
<td>Mitosis detection</td>
</tr>
<tr>
<td>Janowczyk et al. [100,101]</td>
<td>100 × 1000</td>
<td>100</td>
<td>H&amp;E</td>
<td>Breast cancer</td>
<td>Coordinates of lymphocyte</td>
<td>Lymphocyte detection</td>
</tr>
<tr>
<td>Janowczyk et al. [100,101]</td>
<td>1000 × 1000</td>
<td>42</td>
<td>H&amp;E</td>
<td>Breast cancer</td>
<td>Mask for epithelium</td>
<td>Epithelium segmentation</td>
</tr>
<tr>
<td>Janowczyk et al. [100,101]</td>
<td>2000 × 2000</td>
<td>143</td>
<td>H&amp;E</td>
<td>Breast cancer</td>
<td>Mask for nuclei</td>
<td>Nuclear segmentation</td>
</tr>
<tr>
<td>Janowczyk et al. [100,101]</td>
<td>775 × 522</td>
<td>85</td>
<td>H&amp;E</td>
<td>Colorectal cancer</td>
<td>Mask for gland area</td>
<td>Gland segmentation</td>
</tr>
<tr>
<td>Janowczyk et al. [100,101]</td>
<td>50 × 50</td>
<td>277,524</td>
<td>H&amp;E</td>
<td>Breast cancer</td>
<td>None</td>
<td>Tumor detection</td>
</tr>
<tr>
<td>Gertych et al. [22]</td>
<td>1200 × 1200</td>
<td>210</td>
<td>H&amp;E</td>
<td>Prostate cancer</td>
<td>Mask for gland area</td>
<td>Gland segmentation</td>
</tr>
<tr>
<td>Ma et al. [102]</td>
<td>1040 × 1392</td>
<td>81</td>
<td>IHC</td>
<td>Breast cancer</td>
<td></td>
<td>TIL analysis</td>
</tr>
<tr>
<td>Linder et al. [63,103]</td>
<td>93–2372 × 94–2373</td>
<td>1377</td>
<td>IHC</td>
<td>Colorectal cancer</td>
<td>Mask for epithelium and stroma</td>
<td>Segmentation of epithelium and stroma</td>
</tr>
<tr>
<td>Xu et al. [54]</td>
<td>Various size</td>
<td>717</td>
<td>H&amp;E</td>
<td>Colon cancer</td>
<td></td>
<td>Segmentation</td>
</tr>
<tr>
<td>Xu et al. [54]</td>
<td>1280 × 800</td>
<td>300</td>
<td>H&amp;E</td>
<td>Colon cancer</td>
<td>Mask for colon cancer</td>
<td>Segmentation</td>
</tr>
</tbody>
</table>
IDENTIFYING DATASETS

• Problem: NONE ARE DICOM!!!!

• What is DICOM?

  – Digital Imaging and Communications in Medicine (DICOM) is a standard for handling, storing, printing, and transmitting information in medical imaging. It includes a file format definition and a network communications protocol.
PATCH EXTRACTION & TRANSFORMATION

- RGB -> HSV -> Otsu [PMID: 30696894]
- RGB -> HSV -> Otsu -> $0.5 \times t_{Otsu}$ [PMID: 28426134]
- RGB -> Grayscale -> Mean > Threshold [DAPPER]
- RGB -> Lab -> Threshold (L=241)
- RGB -> OD -> SVD [PMID: 28570557]
- [...]
Fig. 1. The same tissue section taken by different microscopic slide scanners. (a) Aperio scanner. (b) Hamamatsu scanner. Image source: ICPR 2014 MITOS-ATYPIA challenge—http://mitos-atypia-14.grand-challenge.org
COLOR NORMALIZATION

VAHADANE et al.: STRUCTURE-PRESERVING COLOR NORMALIZATION AND SPARSE STAIN SEPARATION FOR HISTOLOGICAL IMAGES

Color Normalization

Fig. 3. Flow diagram of the proposed structured preserving color normalization (SPCN).

Per Image Standardization

\[
\frac{x - \hat{x}}{\max(sd, \frac{1}{\sqrt{N}})}
\]

Linearly scales image to have zero mean and unit variance
IMAGE PREPROCESSING

Image reshape, resize, rotate, flip, brightness, contrast, hue, saturation, crop, skew, convert, ...
APPLYING A MODEL

- Many well validated frameworks
**EVALUATING MODELS**

- **sensitivity, recall, hit rate, or true positive rate (TPR)**
  \[ TPR = \frac{TP}{P} = \frac{TP}{TP + FN} = 1 - FNR \]

- **specificity, selectivity or true negative rate (TNR)**
  \[ TNR = \frac{TN}{N} = \frac{TN}{TN + FP} = 1 - FPR \]

- **precision or positive predictive value (PPV)**
  \[ PPV = \frac{TP}{TP + FP} = 1 - PDR \]

- **negative predictive value (NPV)**
  \[ NPV = \frac{TN}{TN + FN} = 1 - FOR \]

- **false rate or false positive rate (FPR)**
  \[ FPR = \frac{FP}{N} = \frac{FP}{FP + TN} = 1 - TNR \]

- **false discovery rate (FDR)**
  \[ FDR = \frac{FP}{FP + TP} = 1 - PPV \]

- **false omission rate (FOR)**
  \[ FOR = \frac{FN}{FN + TN} = 1 - NPV \]

- **accuracy (ACC)**
  \[ ACC = \frac{TP + TN}{P + N} = \frac{TP + TN}{TP + TN + FP + FN} \]

- **F1 score**
  The harmonic mean of precision and sensitivity
  \[ F_1 = 2 \cdot \frac{PPV \cdot TPR}{PPV + TPR} = \frac{2TP}{2TP + FP + FN} \]

- **Matthews correlation coefficient (MCC)**
  \[ MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \]

---

**Dice Coefficient**

\[ DSC = \frac{2|X \cap Y|}{|X| + |Y|} \]

---

**Inter-rater Agreement:**
- **Cohen:** \( N = 2 \)
- **Fleiss:** \( N > 2 \)

\[ k = \frac{p_o - p_e}{1 - p_e} \]

\[ \kappa = \frac{\tilde{P} - \tilde{P}_e}{1 - \tilde{P}_e} \]
ACCURACY CAN BE MISLEADING

Balance training and test sets to prevent "Class Imbalance"

PMID: 25738806
Best Practices in Algorithm Development
DO YOUR HOMEWORK

• Do you need a new algorithm?
• What can you steal use from others?
• Do you need new data, or does some exist?
• Do you have enough data?
• Do you have the right data?
• What model will you use?
• What is your cost function?
• Do you have diversity?
LOOK AT YOUR DATA
IDENTIFY ARTIFACTS

Fig. 3. Artifacts in WSIs. Top: tumor region is outlined with red marker. The arrow indicates a tear possibly formed during the tissue preparation process. Left bottom: blurred image. Right bottom: folded tissue section. The histopathological images are adopted from TCGA [33].

Komura and Ishikawa, 2018

Hashimoto et al, 2012
CHOOSING TRAINING & TESTING SETS

• Balance training an test sets to prevent “Class Imbalance”
  – Use slide-level partitioning instead of patch-level

• Monitor training loss and validation metrics

![Under-fitting](image1.png)  ![Goldilocks-fitting](image2.png)  ![Over-fitting](image3.png)
CLASS IMBALANCE

• Balancing training and test sets to prevent “Class Imbalance”
  – sklearn.model_selection.train_test_split([...], stratify=Labels)
  – sklearn.model_selection.StratifiedKFold(n_splits=3)
USE GOOD CODING PRACTICES

• Write in small, reusable functions

• Write unit tests for your functions
  – Yes, it will eventually save time

• Use version control
Minimum Information about Experiments
MINIMUM INFORMATION

Ensures that the data can be easily verified, analyzed and clearly interpreted by the wider scientific community. Keeping with these recommendations also facilitates the foundation of structuralized databases, public repositories and development of data analysis tools.

1 MI Standards
1.1 MIAPPE, Minimum Information About a Plant Phenotyping Experiment
1.2 MIAME, gene expression microarray
1.3 MINI: Minimum Information about a Neuroscience Investigation
   1.3.1 MINI: Electrophysiology
1.4 MIARE, RNAi experiment
1.5 MIACA, cell based assay
1.6 MIAP, proteomic experiments
1.7 MIMx, molecular interactions
1.8 MIAPAR, protein affinity reagents
1.9 MIABE, bioactive entities
1.10 MIINS, genome/metagenome sequences
1.11 MIFlowCyt, flow cytometry
1.12 Minimum Information about a Flow Cytometry Experiment
1.13 MISFISHIE, In Situ Hybridization and Immunohistochemistry Experiments
1.14 MIAPA, Phylogenetic Analysis
1.15 MIRAGE, Glycomics
1.16 MIQ, ORF
1.17 MIAMET, METabolomics experiment
1.18 MIAFGE, Functional Genomics Experiment
1.19 MIRIAM, Minimum Information Required in the Annotation of Models
1.20 MIASE, Minimum Information About a Simulation Experiment
1.21 CIMR, Core Information for Metabolomics Reporting

Commentary | Published: 01 December 2001

Minimum information about a microarray experiment (MIAME)—toward standards for microarray data
MIA MODELS

MIA-WSI

• Minimum information about Whole Slide Image Analysis
  – Stain Type
  – Magnification Levels
  – DICOM Tags

MIA-MODEL

• Minimum information about Model Development
  – Sample list part of train/val/test
  – Number of epochs
  – Model architecture
  – Weights
  – Loss/Accuracy Curves
  – Baseline for comparison
Opportunities for
Shared Code/APIs
STANDARD APIs

- `downloadData(setName, out_dir)`
- `excludeAritfacts(image)`
- `hasEnoughTissue(image, threshold=0.85)`
- `passesQC(image)`
- `normalizeImage(image, method="Otsu", value_norm=True)`
- `preProcessImage(image, methods=['flip', 'rotate', 'skew', ...])`
- `trainModel(image, label, model)`
- `runModel(image, model)`
- `evaluateModel(images, labels, model)`
PIPELINES & MODULES

ML_archetype/
  └── datasets
      ├── dataset_factory.py
      └── dataset_utils.py
  └── deployment
      └── model_deploy.py
  └── docs
      └── authors.rst
      └── Makefile
      └── usage.rst
  └── evaluate_model.py
  └── LICENSE
  └── Makefile
  └── nets
      └── nets_factory.py
  └── preprocessing
      └── preprocessing_factory.py
  └── README.md
  └── requirements.txt
  └── tests
      └── test_partA.py
      └── test_partB.py
  └── train_model.py
  └── tutorials
      └── tutorial.ipynb
  └── utils
      └── utilities.py
GROUP DISCUSSION

COULD/SHOULD API OR HIMA ESTABLISH AND MAINTAIN A PRESENCE ON GITHUB?

Who would maintain it?
How would it be managed?
What would be in scope?
What are the criteria for a Pull Request?
LEARN FROM OTHERS!
THANK YOU
SAM Questions
SAM Question #1

Which is the standard for the representation, storage, and communication of medical images and related information:

- WSI
- DICOM
- TIFF
- SVS

• Answer: DICOM

• TIFF and SVS are common file formats, but not the standard

• WSI is a generic term for any image format
Which of the following is a free online repository that hosts whole slide images?

- TCIA
- GitHub
- dbGaP
- Aperio

• Answer: TCIA
  - The Cancer Imaging Atlas

• GitHub stores exchangeable code, not data
• dbGaP controls access to phenotypic data
• Aperio is a type of WSI scanner

• See also the DPA website:
  - https://digitalpathologyassociation.org/whole-slide-imaging-repository
In machine learning classification, what is meant by the term “class imbalance”? 

- The training and testing sets are different sizes
- Samples that are harder to train are only in the test set
- One of the classes you are trying to predict is over- or underrepresented
- There is a lack of diversity in the training set

• Answer: One of the classes you are trying to predict is over- or underrepresented

• Has nothing to do with the difference in size or the complexity of the task between training and test sets
What is unit testing?

- Testing that a function works as expected
- Testing a series of connected functions works as expected
- Testing the amount of memory or compute needed to function
- Ensuring that the most important functions in an application work

• Answer: **Testing that a function works as expected**

• All other tests are necessary, but have different purposes. Unit testing is focused on very small functions with a limited scope. Good automated unit tests will save time, enable reuse of code, and make long-term support possible.