AUTOMATED PATHOLOGY DETECTION SYSTEM DESIGN USING MEDICAL IMAGES: AN AZURE MACHINE LEARNING RESEARCH

Presented By:
Sohini Roy Chowdhury
Assistant Professor, Department of Electrical Engineering
University of Washington, Bothell, WA, USA
roych@uw.edu
OUTLINE

- The Problem: Automated Pathology Detection using Medical Images
- Methodology
- Motivation for Microsoft Azure Machine Learning Studio (MAMLS) Platform
- MAMLS Performance
- Results
- Conclusions
MEDICAL IMAGE DIAGNOSTICS: THE CHALLENGES

- Background noise.
- Imaging artifacts.
- Variations in imaging systems.
- Telemedicine.
- Biometrics.

Retinal image analysis system.

Semi-automated ultrasound analysis system.

Source: http://en.wikipedia.org/wiki/Medical_imaging
RETINAL FUNDUS IMAGE ANALYSIS

- **Fundus**: Image of the interior surface of the eye, opposite the lens.

- Pathologies causing acquired blindness:
  - Non-Proliferative Diabetic Retinopathy (NPDR).
  - Proliferative Diabetic Retinopathy (PDR).
  - Glaucoma.
  - Age-related Macular Degeneration (AMD).
  - Retinal Vein Occlusion (RVO).

Source: [Online]
http://www.ucdenver.edu/academics/colleges/medicalschool/departments/Ophthalmology/research/labresearch/Faculty/PublishingImages/Figure%202%20V1.jpg
METHODOLOGY: INFORMATION EXTRACTION FROM MEDICAL IMAGES

- Cognitive symptoms of pathology \[\rightarrow\] Detection metrics.

- The tools:
  - Domain knowledge - For intuitive understanding.
  - Image Processing - For segmenting Regions of Interest (ROI).
  - Machine Learning - For decision making.

- Outcome:
  - Automated algorithms for pathology detection.
  - Semi-automated algorithms to aid treatment.
  - Reliable Telemedicine.
DIABETIC RETINOPATHY ANALYSIS USING MACHINE LEARNING: THE DREAM SYSTEM

Stage 1: Background Detection

Stage 2: Bright Lesion Classification

Stage 3: Red Lesion Classification

Foreground segmentation  Lesion Classification: Step 1  Lesion Classification: Step 2
MOTIVATION FOR CLOUD COMPUTING

- Existing challenges:
  - Classification tasks incur high computational time complexities.
  - Identification of pathology specific medical informatics.
  - Scalable metrics/methods for large populations for screening tasks.

- Motivation for using Microsoft Azure Machine Learning Studio (MAMLS):
  - Selection of algorithms and packages in R.
  - Flow-chart based user interface aids end-users.

- Key Contributions:
  - Automate classification tasks within MAMLS.
  - Benchmark MAMLS platform.
THE MAMLs PLATFORM:

- Data set
THE MAMLS PLATFORM:

- Data set
- Split
THE MAMLS PLATFORM:

- Data set
- Split
- Train
THE MAMLS PLATFORM:

- Data set
- Split
- Train
- Test
MAMLS MACHINE LEARNING PROCESS

- Data set
- Split
- Train
- Test
- Evaluate
THE MAMLS PLATFORM:

- Data set
- Split
- Train
- Test
- Evaluate
  - Visualize
MAMLs Performance

- Public Data Sets:
  - Telescope\(^2\)
  - Wisconsin Breast Cancer\(^3\)

- Local Data Set:
  - Non Proliferative Diabetic Retinopathy lesions: DIARETDB1
    - Available at: [https://sites.google.com/a/uw.edu/src/useful-links](https://sites.google.com/a/uw.edu/src/useful-links)
LOCAL DATA SET – DIARETDB1

- 4 lesions for Diabetic Retinopathy -> 6 classes
  - Bright Lesion: False Positive, 0
  - Bright Lesion: Hard Exudates, 1
  - Bright Lesion: Soft Exudates, 2
  - Red Lesion: False Positive, 3
  - Red Lesion: Micro-Aneurysm, 4
  - Red Lesion: Hemorrhage, 5

![Bar Chart showing frequency distribution of Col67 with rows 15945 and columns 67]
Dataset: User Input

Decision 1: Binary or Multi-class

Decision 2: Binary

Decision 2: Multi-class

Decision 3: Multi-class or Hierarchical Binary

Decision 4: Hierarchical Binary

Result: Flow Output

PROPOSED FLOW
PROPOSED FLOW

Dataset: User Input

Decision 1: Binary or Multi-class

Decision 2: Binary

Decision 2: Multi-class

Decision 3: Multi-class or Hierarchical Binary

Decision 4: Hierarchical Binary

Result: Flow Output
PROPOSED FLOW

Dataset: User Input

Decision 1: Binary or Multi-class

Decision 2: Binary

Decision 2: Multi-class

Decision 3: Multi-class or Hierarchical Binary

Decision 4: Hierarchical Binary

Result: Flow Output
PROPOSED FLOW

Dataset: User Input

Decision 1: Binary or Multi-class

Decision 2: Binary

Decision 2: Multi-class

Decision 3: Multi-class or Hierarchical Binary

Decision 4: Hierarchical Binary

Result: Flow Output
PROPOSED FLOW

Dataset: User Input → Decision 1: Binary or Multi-class → Decision 2: Binary or Multi-class → Decision 3: Multi-class or Hierarchical Binary → Decision 4: Hierarchical Binary → Result: Flow Output
PROPOSED FLOW

Dataset: User Input

Decision 1: Binary or Multi-class

Decision 2: Multi-class

Decision 2: Binary

Decision 3: Multi-class or Hierarchical Binary

Decision 4: Hierarchical Binary

Result: Flow Output
RESULTS – TELESCOPE²

Generalized Flow - Telescope ➔ Execute R Script ➔ Result Dataset

<table>
<thead>
<tr>
<th>rows</th>
<th>columns</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Acc</th>
<th>Error</th>
<th>Sen</th>
<th>Spec</th>
<th>K-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-Nearest Neighbor</td>
<td>0.8137</td>
<td>0.1863</td>
<td>0.5491</td>
<td>0.958</td>
<td>25</td>
</tr>
</tbody>
</table>
RESULTS – WISCONSIN BREAST CANCER

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Acc</th>
<th>Error</th>
<th>Sen</th>
<th>Spec</th>
<th>K-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-Nearest Neighbor</td>
<td>0.9603</td>
<td>0.0397</td>
<td>0.9586</td>
<td>0.9613</td>
<td>5</td>
</tr>
</tbody>
</table>

Rows: 1
Columns: 6
RESULTS – DIARETDB1

DIARETDB1 hierarchical classification ROC curves. (a) level 1. (b) level 2. (c) level 3. (d) level 4. (e) level 5.
CONCLUSIONS AND FUTURE WORK

- MAMLs
  - Binary, Multi-class, and Hierarchical Binary classification tasks.

- Proposed Generalized Flow
  - Fine-tuned classification models.
  - Decision-making R scripts.

Future Work

- Dimensionality reduction.
- Output visualization.
- Regression capability.
- Benchmark additional public and local data sets.
REFERENCES


THANK YOU

- Questions??