

Preliminary Design Review Presentation

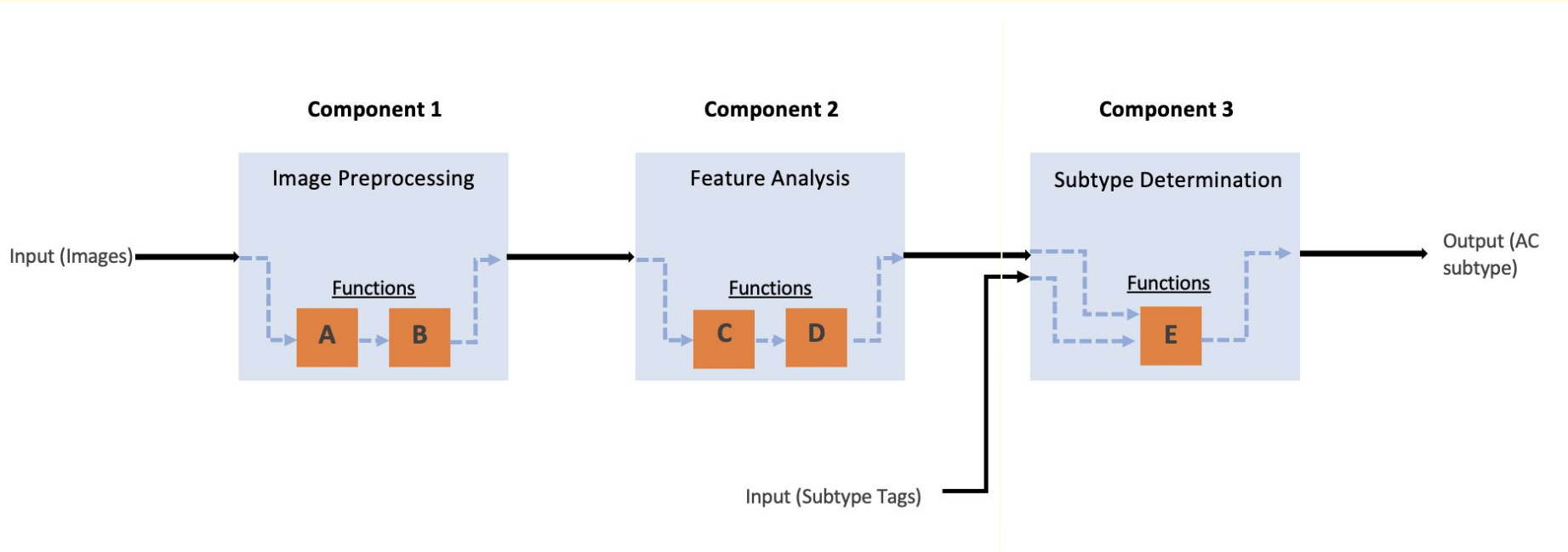
Machine Learning Algorithm for the
Classification of Appendiceal Cancer

Capstone Fall 2020



WAKE FOREST
UNIVERSITY

- Updated system requirements to reflect functions that are independent of the solutions generated
 - Explored other solutions in addition to the deep learning algorithm
 - Researched image processing and augmentation software such as MATLAB, QuPath, and Visiopharm and specific capabilities of each.
 - Developed an general overview of 2 possible solutions
 - Discussed pros vs cons of each solution against a criteria
 - Identified some key features of the subtypes
-



A. Read pixel information into data structures

B. Modify the images to account for real-world differences in H/E stained slides

C. Down sample pixel-level data until only desirable features are represented

D. Categorize features based on similarity and diagnostic relevance.

E. Interpret the relationship between features and subtype tags.

Component 1 (Image Preprocessing)

Slide Analysis
Application (QuPath)

MATLAB Image
Processing Toolbox

Open-source Data
Augmentation Packages

Component 2 (Feature Analysis)

MATLAB Image
Processing Toolbox

Convolutional Network

Component 3 (Subtype Determination)

Machine Learning

Statistical Analysis

Component 1 (Image Preprocessing)

Slide Analysis
Application (QuPath)

MATLAB Image
Processing Toolbox

Open-source Data
Augmentation Packages

Component 2 (Feature Analysis)

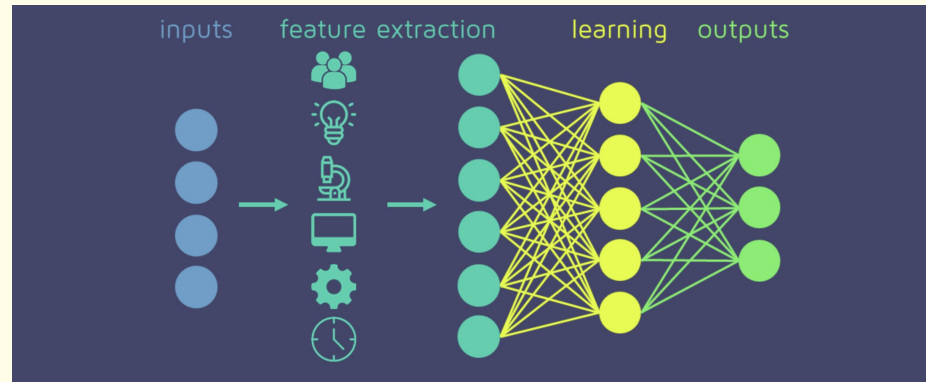
MATLAB Image
Processing Toolbox

Convolutional Network

Component 3 (Subtype Determination)

Machine Learning

Statistical Analysis



- **Solution A: Features Image Processing Model will Extract**

Mucin as a percentage of tumor area

Epithelial tissue presence in mucin pools

Shape of tumor perimeter

Cytologic atypia

Degree of fibrosis in stroma

Local glandular development

Signet ring cells as a percentage of tumor area

Nuclear stratification of epithelium

Degree of lymph node inflammation

Mean distance between tumors

Median tumor diameter

Percentage of total image that is tumor

Component 1 (Image Preprocessing)

Slide Analysis
Application (QuPath)

MATLAB Image
Processing Toolbox

Open-source Data
Augmentation Packages

Component 2 (Feature Analysis)

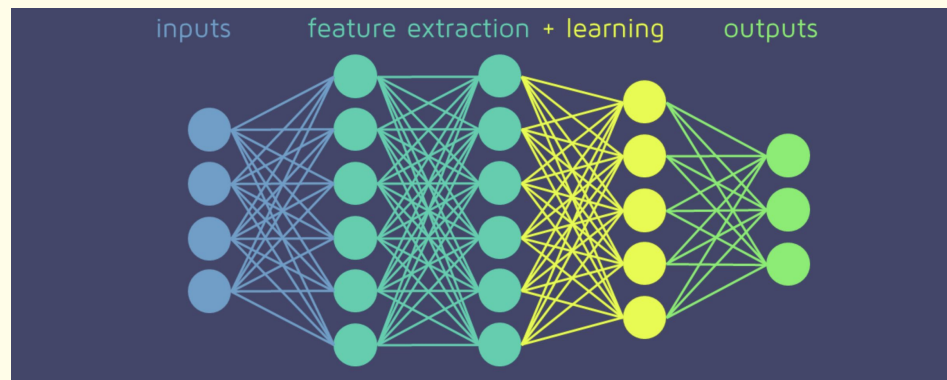
MATLAB Image
Processing Toolbox

Convolutional Network

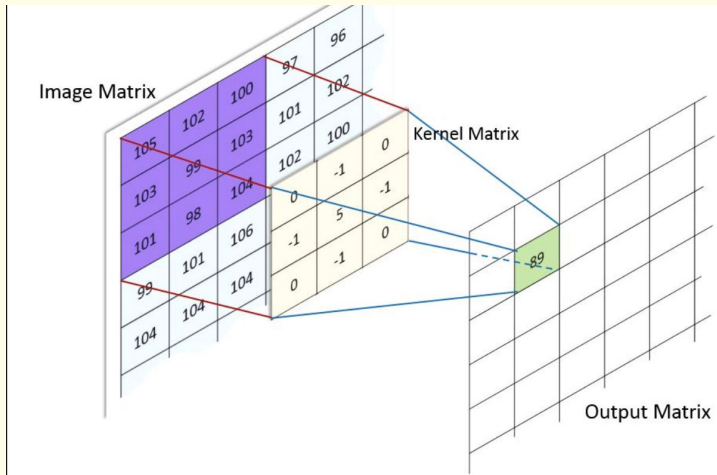
Component 3 (Subtype Determination)

Machine Learning

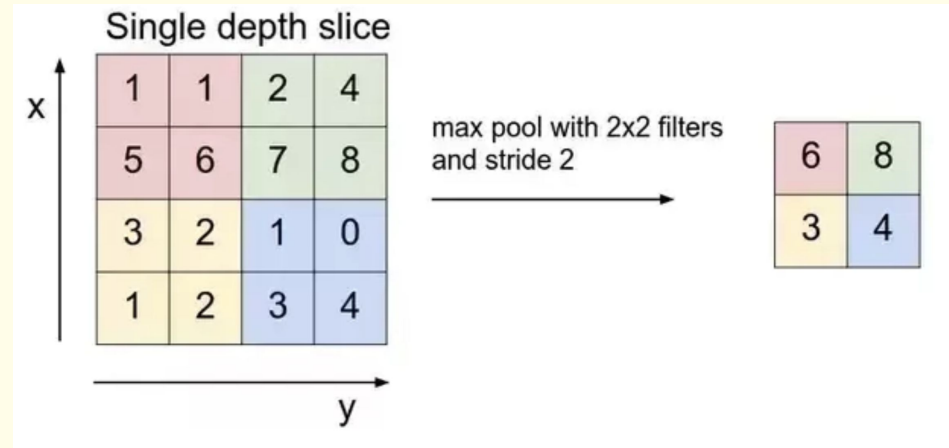
Statistical Analysis



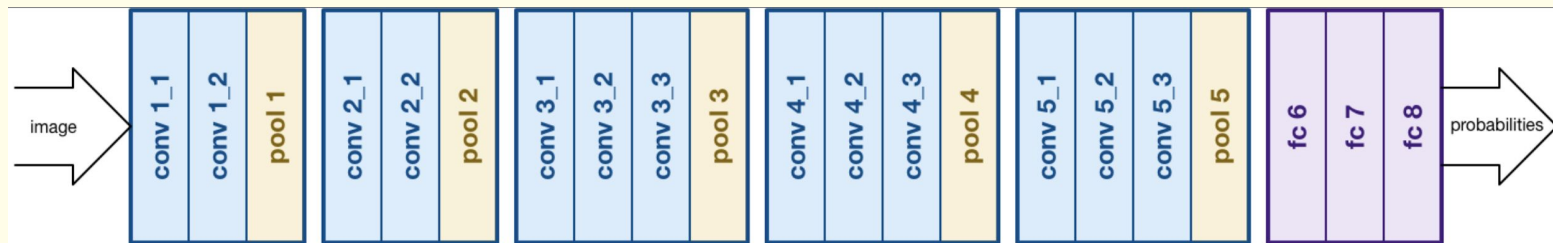
Convolution:



Max Pooling:



- Various combinations of these layers are used in sequence depending on problem
- Convolutional filters extract low, medium, and high-level features



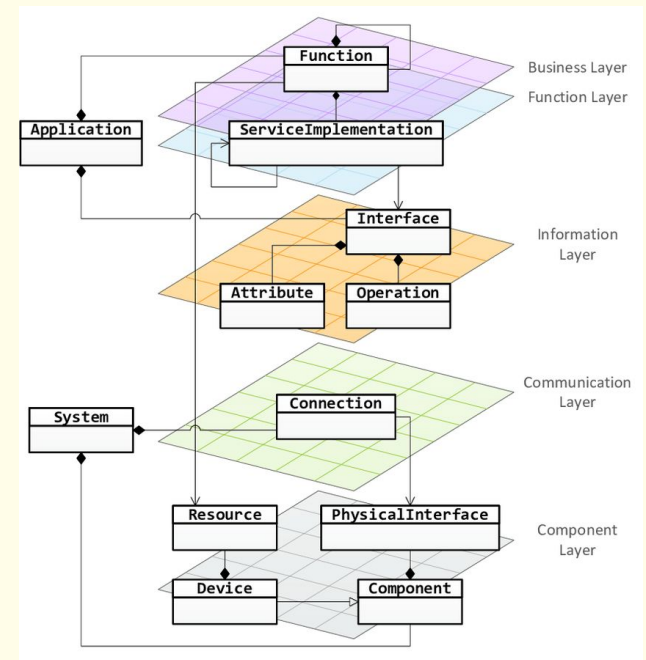
- **Solution A: Image Processing Algorithm + machine learning**
 - Requires fewer images.
 - More control over extracted features.
 - Inherits problems from the Pathologist's approach.
 - Allows flexibility for additional stains to be added in the future
 - **Solution B: CNN**
 - Better performance given large training set.
 - Discovers and extracts its own features.
 - Requires 10,000+ images.
-

	Amount of Images	Timeline	Accuracy	Accessibility and Price	Ease of Development	Future Development	Total
Criteria Weight	5	4	6	3	3	3	
Solution A MATLAB image processing + machine learning	7	7	7	7	6	7	165
Solution B Deep learning (CNN)	5	5	9	6	7	6	156

- **Final Concept Selection**
 - Choose either Solution A or Solution B
- **Iterate on System Architecture Diagram**

- Include more detailed information regarding the specific techniques we will implement to perform preprocessing, feature extraction, etc.

- **Training, developing, and debugging software solution**
- **Function, Fit, Form, & Finish**
- **Create artifacts to communicate pieces of the program**
 - Pseudocode, flow charts, use cases, Unified Modeling Language (UML), class diagrams, risk assessments, source code



Questions

- What are your thoughts on the system architectures of our solutions?
 - Are there other system architectures, tools, or platforms we haven't considered?
-