A DEEP LEARNING BASED PIPELINE FOR METASTATIC BREAST CANCER CLASSIFICATION FROM WHOLE SLIDE IMAGES (WSI)

Presenter: Arjun Vekariya

Supervising Professor: Dr. Junzhou Huang

University of Texas at Arlington

Outline

- **Introduction**
- **Classical methods** 
- **Deep learning-based pipeline to detect metastatic breast cancer**
  - Region of Interest (ROI) detection with Image processing.
  - Prepare training data: Extract Positive & Negative tiles from ROI
  - Train Deep ConvNet for tile-based classification
  - Building tumor probability heat-maps using trained model
  - Post-processing of heat-maps for slide-based classification
- **Experiments**
- **Conclusion & Future work**

Problem & Motivation

~ 10^6 px

~ 10^6 px

500 px

375 px

- Problems
  - Very large Gigapixel images **(10^6 x 10^6)**
  - Hard to scan whole image manually
  - Need rapid, cheaper and precise primary diagnosis method
  - Classical computing methods are not robust， uses low level image analysis tasks； (e.g., color normalization, nuclear segmentation, and feature extraction)
  - No Deep Learning base method for Gigapixel image classification
- Motivation
  - Recently, deep learning-based approaches shown promises for applications in Pathology[1]

1. D. C. Cires¸an, A. Giusti, L. M. Gambardella, and J. Schmidhuber. Mitosis detection in breast cancer histology images with deep neural networks.

Whole Slide Image (WSI) Classification

- What?
  - Distinguish tumor positive (Cancer) slides from negatives
- Why?
  - To determine presence and severity of cancer

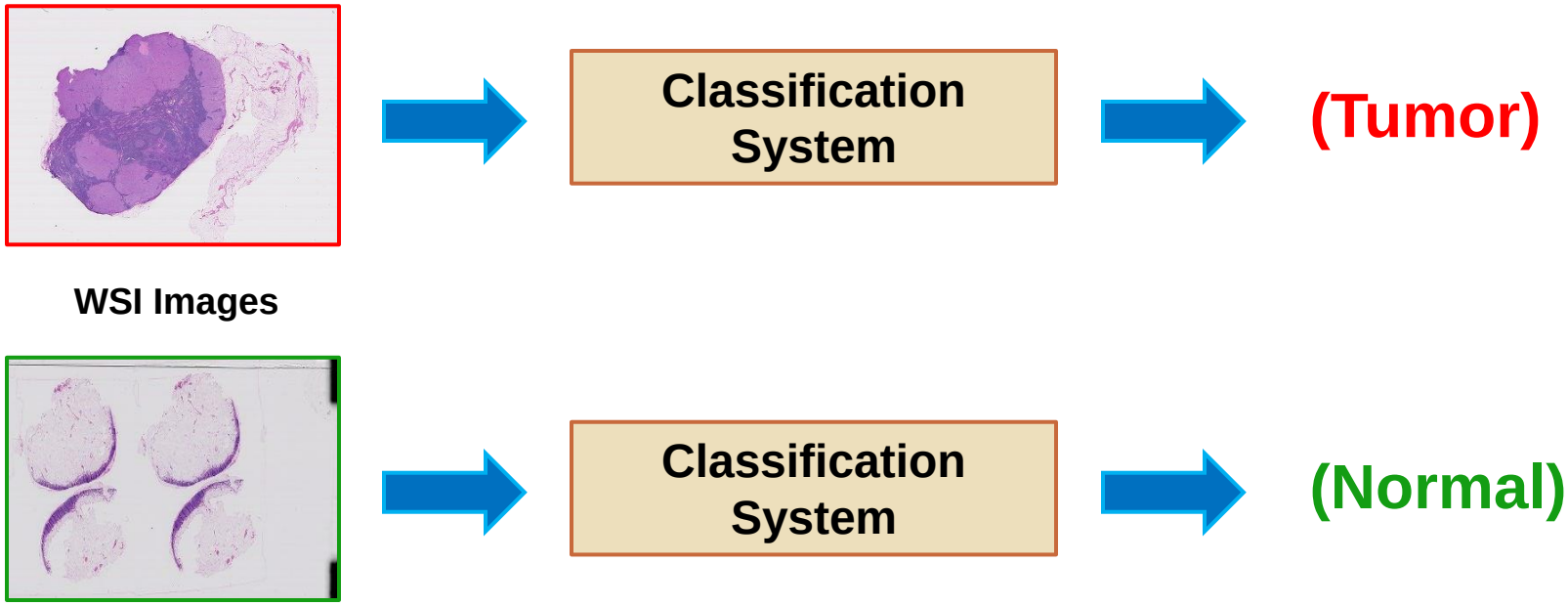


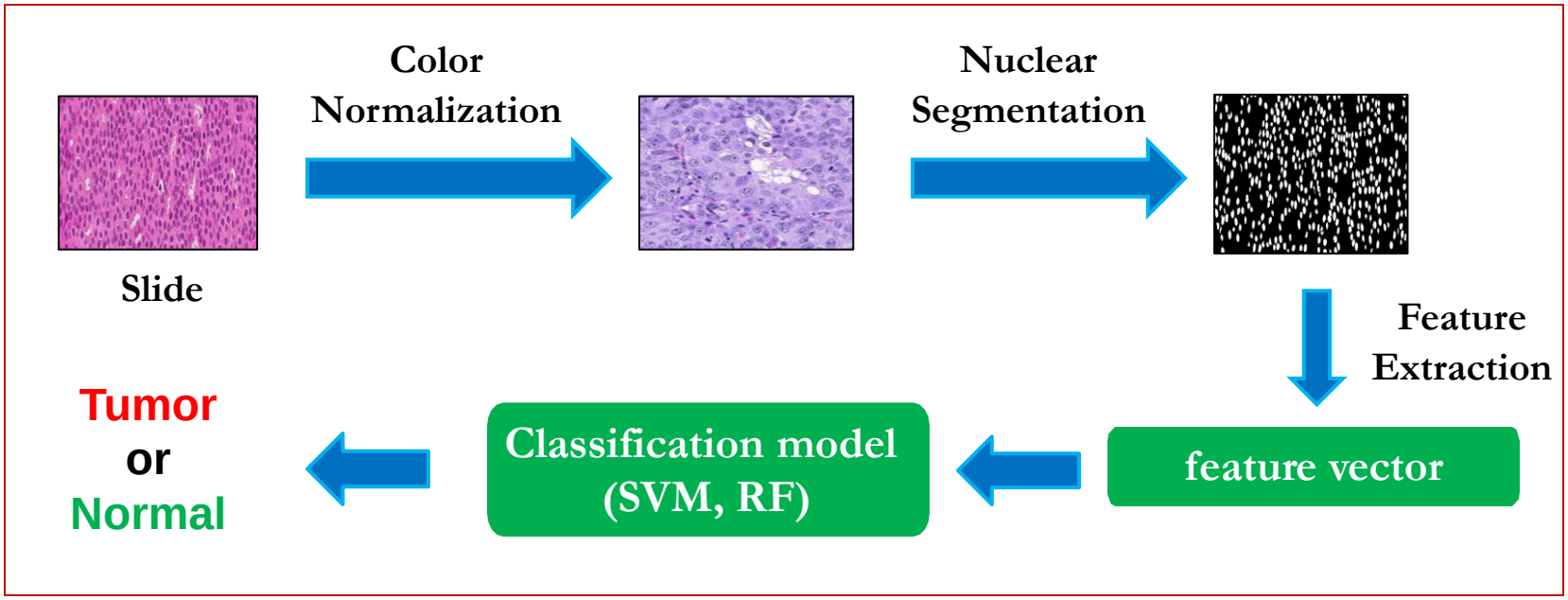
Figure: WSI classification example

Goal of Thesis

- Develop deep learning-based classification pipeline for detection of cancer metastases from Gigapixel whole slide images of breast sentinel lymph node
- Evaluate its effectiveness by performing extensive experiments on real life breast cancer data-set available as part of Camelyon'16 grand challenge

Classical methods

- Focused primarily on low level image analysis tasks
  - Color normalization
  - Nuclear Segmentation
  - Feature extraction
- Architecture



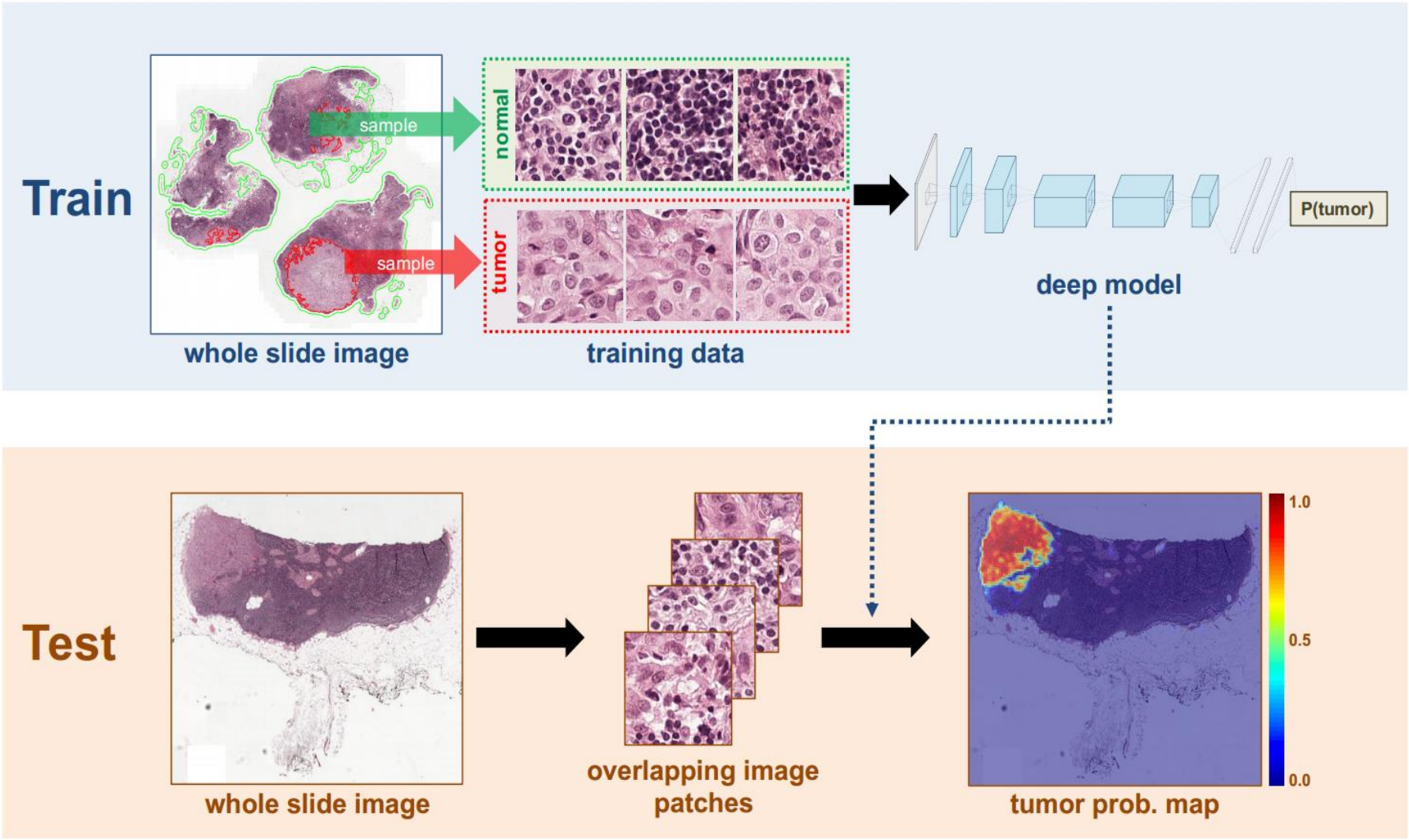
Classical methods (cont.)

- CHALLENGES
  - Relies on a-priori information. Example : shape, size of a cell
  - Handcrafted features do not perform well
  - Requires to set several manual parameters thus proves burdensome in practice
  - Not generalized; fails when applied on images other than what it originally developed for

Our approach : OUTLINE

- In this thesis we try to overcome the challenges of classical methods by developing state-of-the-art deep learning based classification pipeline for detection of cancer metastases.
- Pipeline consists of five stages:
  - Region of Interest (ROI) detection with Image processing.
  - Construct training data: Extract Positive & Negative tiles from ROI
  - Train Deep ConvNet for tile-based classification
  - Building tumor probability heat-maps using trained model
  - Post-processing on heat-maps for slide-based classification

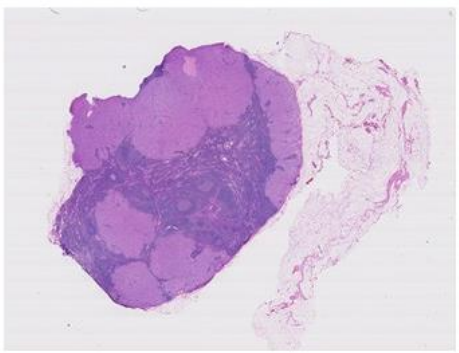
Cancer metastases detection framework[2]



2. D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck, "Deep Learning for Identifying Metastatic Breast Cancer," arXiv preprint arXiv:1606.05718

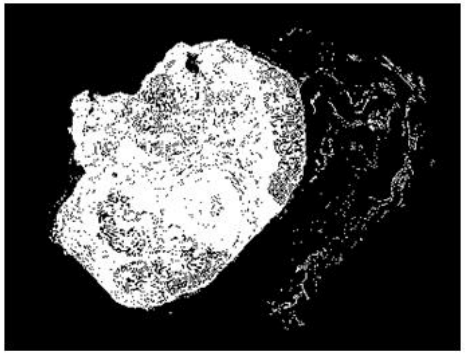
ROI detection with Image processing

- **What?** - identifying tissue within the WSI and exclude background white space
- **Why?** - reduce computation time and focus analysis on regions of the slide most likely to contain cancer metastasis



RGB

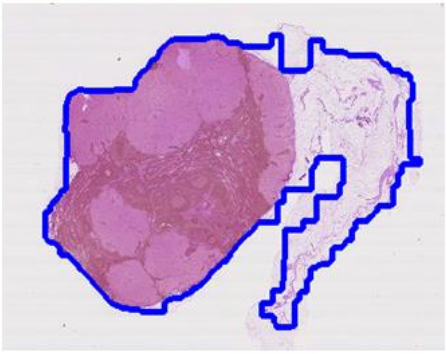
HSV (hue, saturation and value)



Filtered mask

Closing

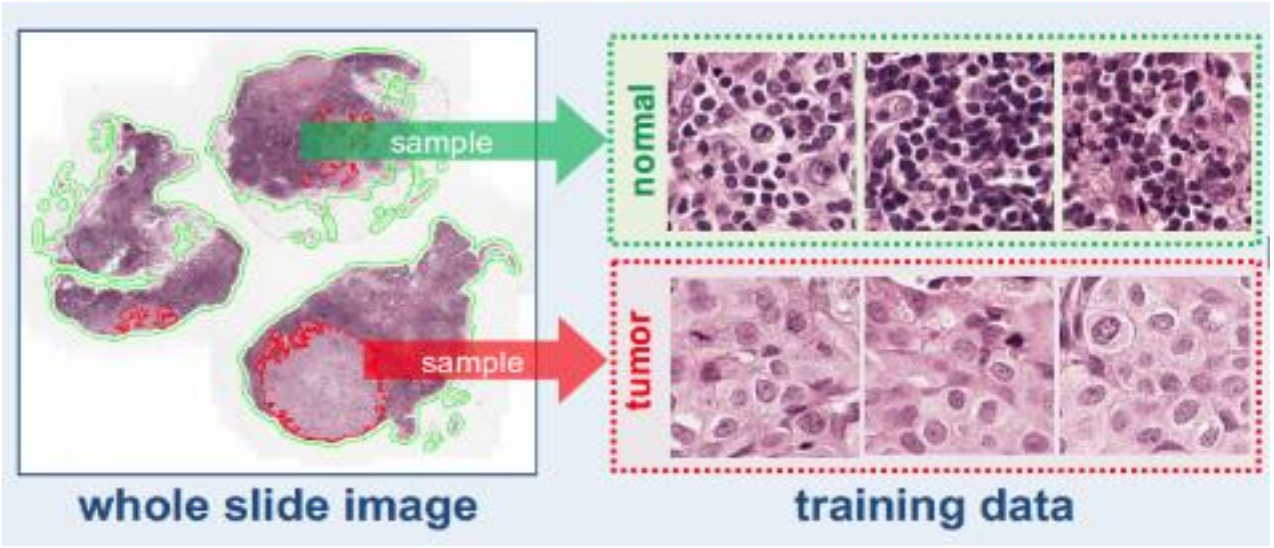
Opening



Contour (ROI)

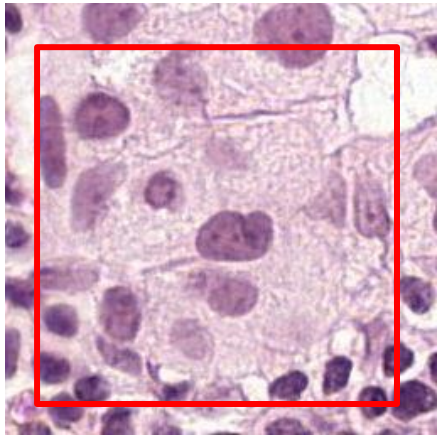
Training set construction

- Randomly extract patches **(256 x 256)** from tissue region (ROI)
  - Tumor slide : ~**1k** positive and ~**1k** negative from each slide
  - Normal slide: ~**1k** negative from each slide
  - Total ~**250k** training patches, ~**140k** normal and ~**110k** positive
- Patches are extracted from **level-0** (highest-**40x** magnification) of each WSI



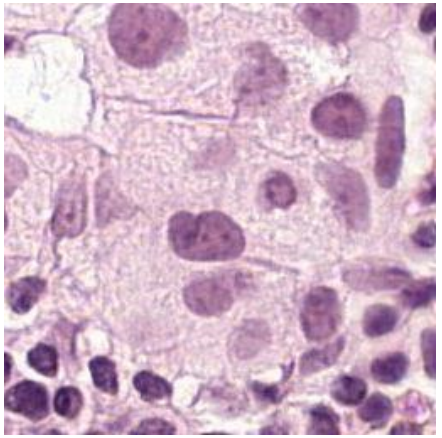
Data augmentation

- Randomly crop a 224 x 224 sub-region and flip patches horizontally



Crop

or



Horizontal flip

Deep ConvNet for tile-based classification

- What in ConvNet?
  - ConvNet is a biologically inspired form of artificial neural network.
  - Stack of layers.
  - There are three main types of layers to build ConvNet:
    - Convolutional Layer + Non-Linearity (ReLU)
    - Pooling Layer
    - Fully-Connected Layer

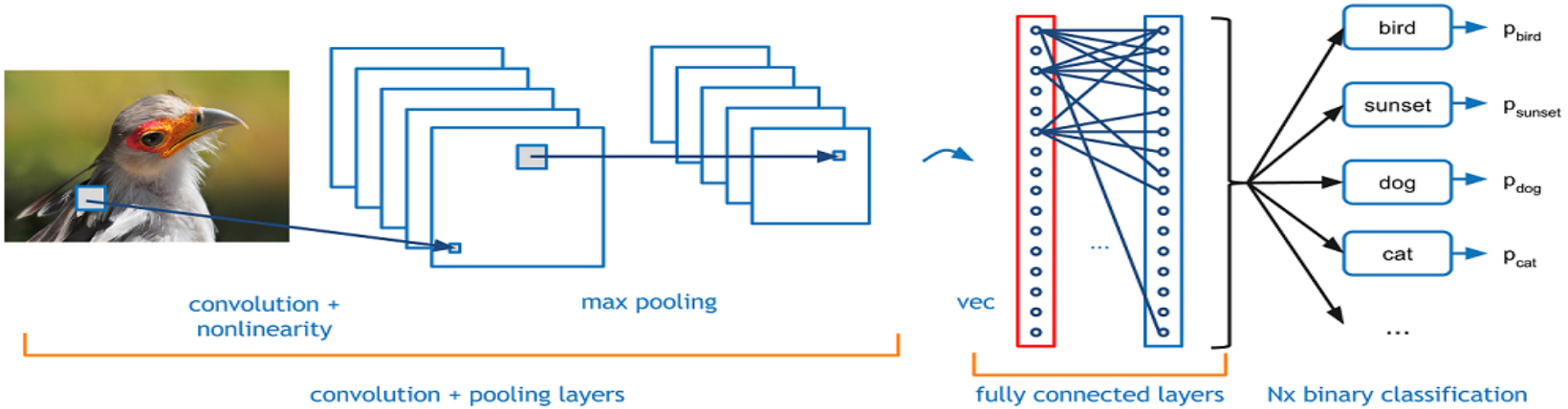
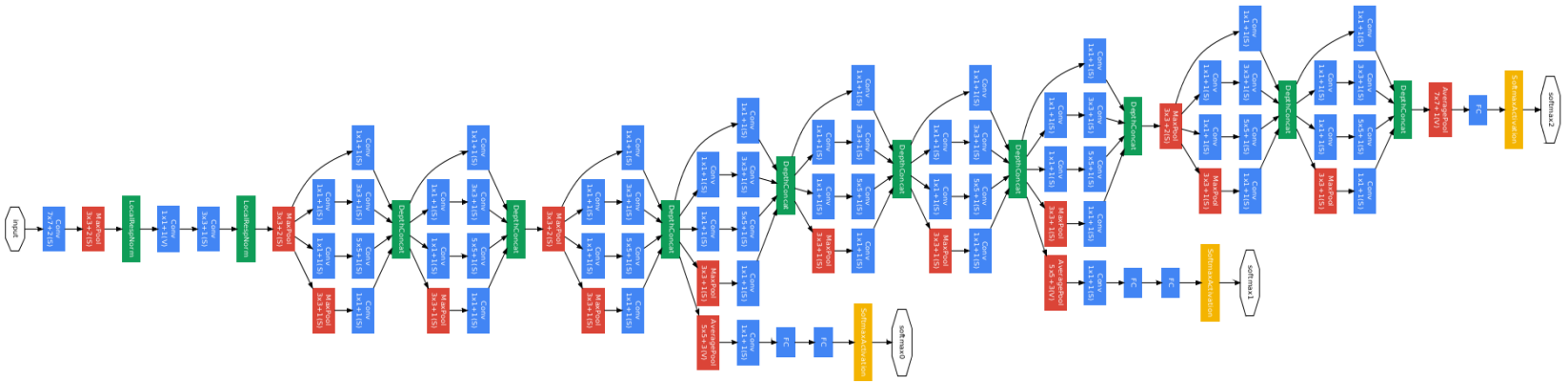


Figure 6: Illustration of CNN

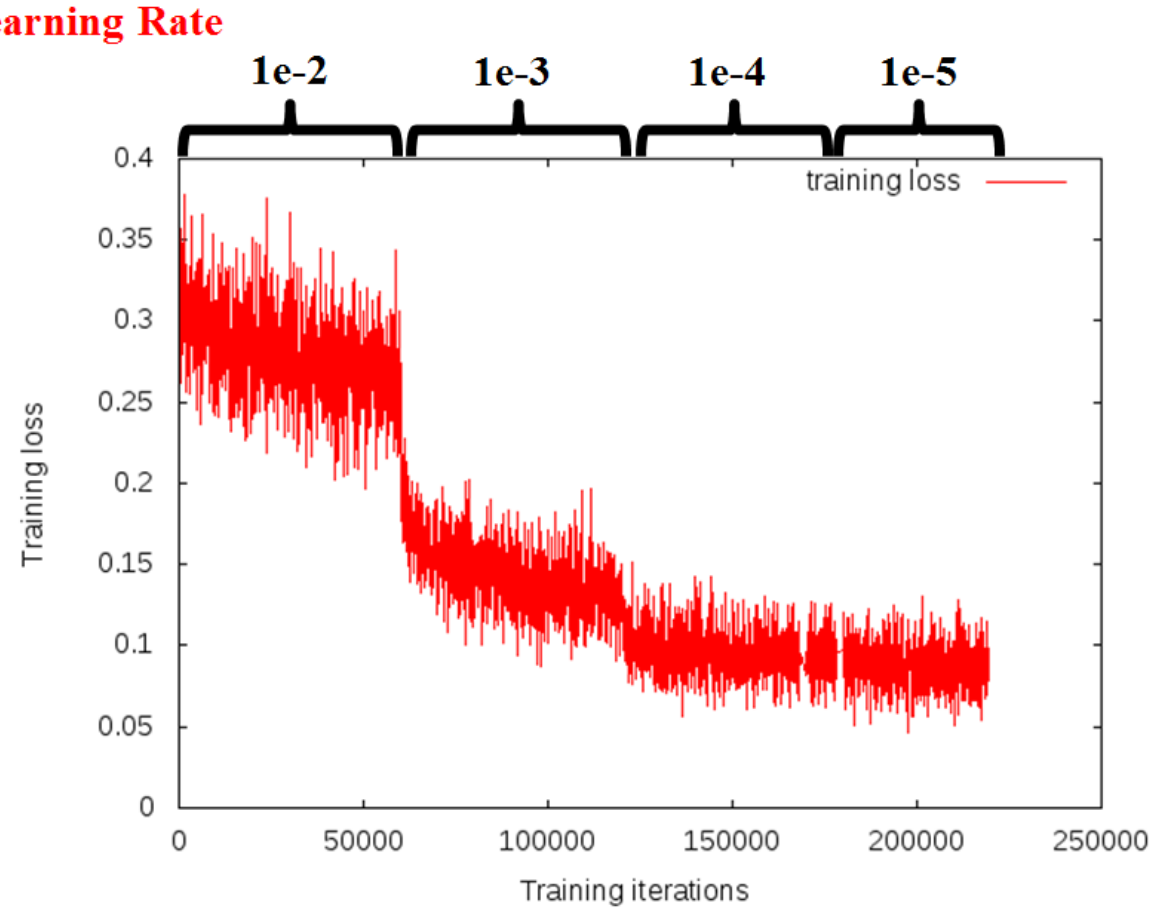
GoogLeNet

- GooLeNet (Szeged et al. ILSCV 2014)
  - 27 layers in total
  - ~6 million parameters
  - three loss layers
  - Source: Christia Szeged et al. Going Deeper with Convolutions
- GooLeNet Architecture:



Training GoogLeNet

- Deep model is trained from scratch using **mini-batch SGD**
- Batch size: **32**



Environment:

- GPU:

  2 x 12 GB NVidia K40
- CPU:

  3.4GHz Intel core i7

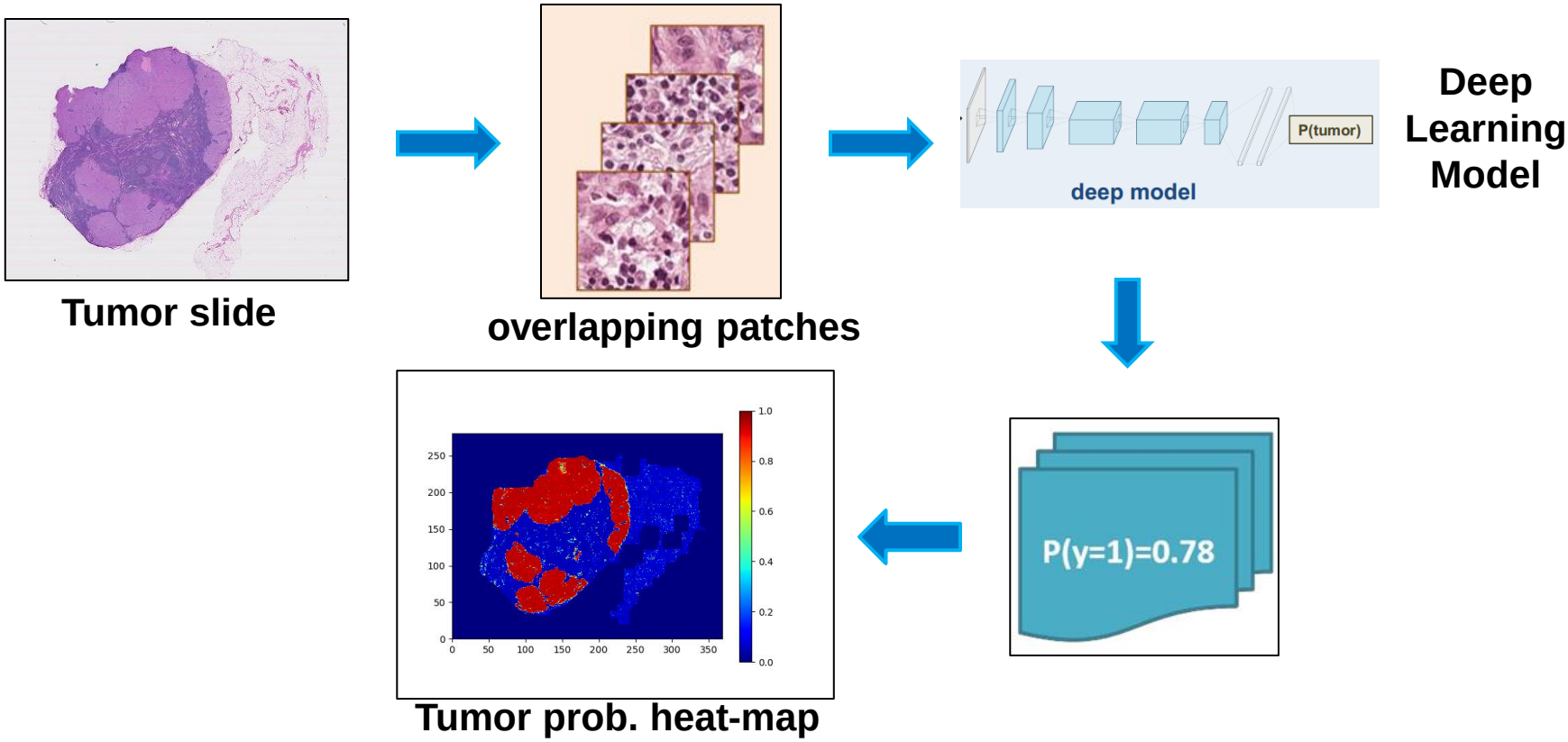
  4770
- HDD:

  7 TB
- RAM:

  16 GB DDR4

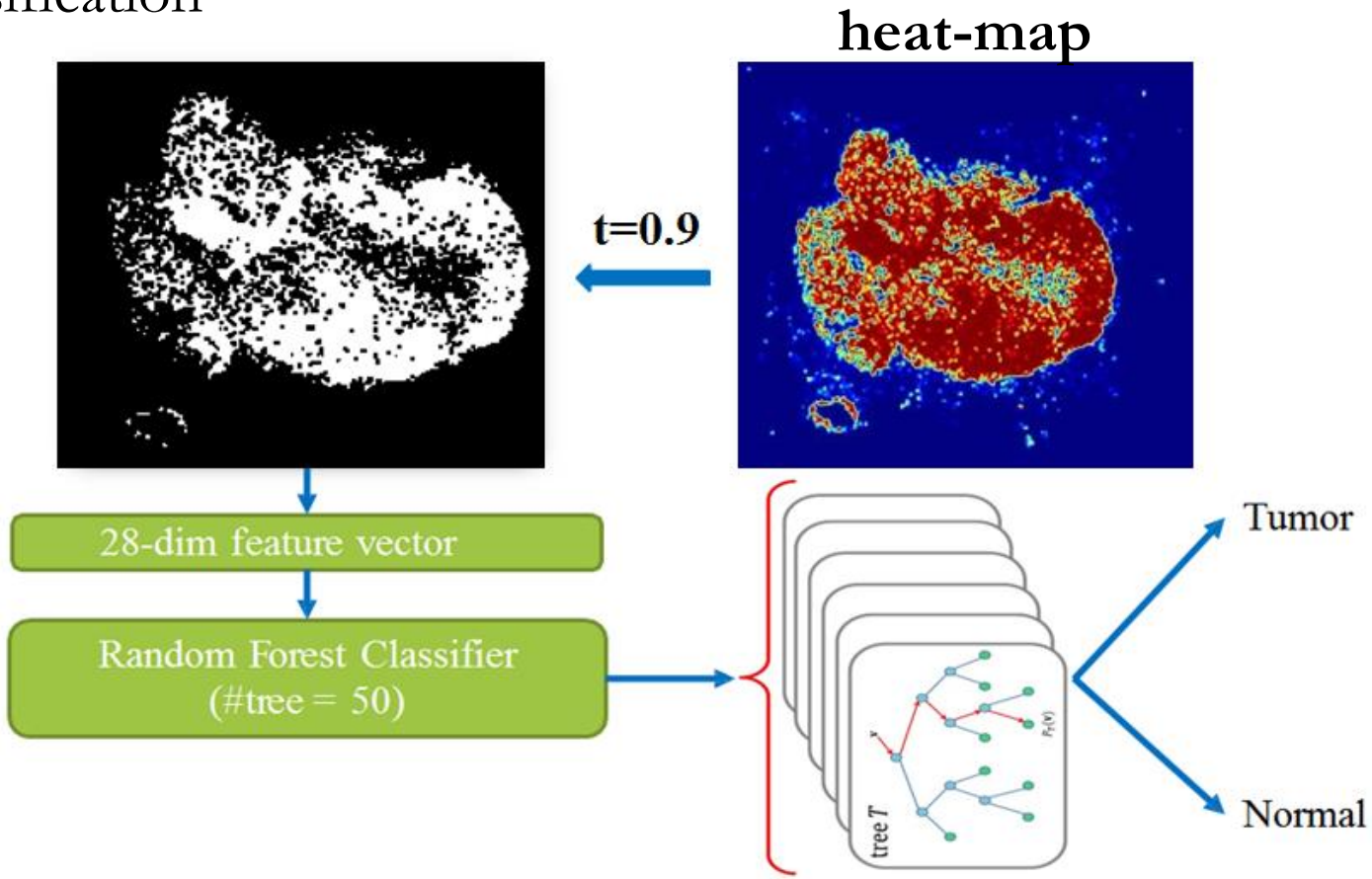
Building tumor probability heat-maps

- Extract patches from ROIs of each WSI, ~ **7.6M** total patches
- Use trained Deep CNN model to build heat-map for each WSI
- In heat-map, each pixel contains a value between 0 and 1, indicating the probability that the pixel contains tumor

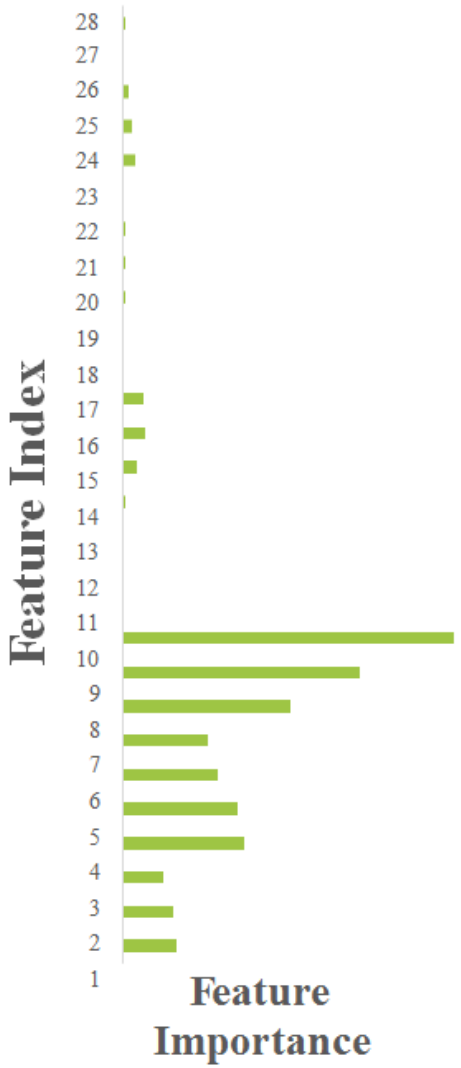


Post-processing on tumor probability heat-maps

- Extract higher level features from tumor probability heat-maps
- Using extracted features, train binary classifier for slide based classification



Feature Importance



- **Top 5** important features, computed using the **"regionprops"** function in skimage. **t** is the threshold value
  - **Feature 11:** given t=0.9, mean area of tumor regions
  - **Feature 10:** given t=0.5, the longest axis in the largest tumor region
  - **Feature 09:** given t=0.5, ratio of pixels in the region to pixels in the total bounding box ("extent")
  - **Feature 05:** given t=0.9, eccentricity of the ellipse that has the same second-moments as the region. ("eccentricity")
  - **Feature 06:** given t=0.9, ratio of tumor region to the tissue region

Complete list of features

- Ratio of tumor region to the tissue region
- The longest axis in the largest tumor region
- Total number of pixels with probability **> 0.90**
- **Tumor area**
  - max, mean, variance, skewness, and kurtosis of tumor area
- **Tumor perimeter**
  - max, mean, variance, skewness, and kurtosis of tumor perimeter
- **Eccentricity** (of ellipse having same second-moments as region)
  - max, mean, variance, skewness, and kurtosis of eccentricity
- **Extent** (ratio of pixels in the region to pixels in the total bounding box)
  - max, mean, variance, skewness, and kurtosis of extent
- **Solidity**
  - max, mean, variance, skewness, and kurtosis of solidity

Experiments: DATASET

- Camelyon'16[1] grand challenge dataset
  - 400 slides in total
  - Train (270)
    - 110 tumor (positive) slides with ground truth
    - 160 normal (negative) slides
  - Test (130)
    - 130 unlabeled slides

1. https://camelyon16.grand-challenge.org/

Experiments: SETUP

- Hardware Configurations:
  - CPU: 3.4GHz Intel core i7 4770
  - RAM： 12 GB DDR4

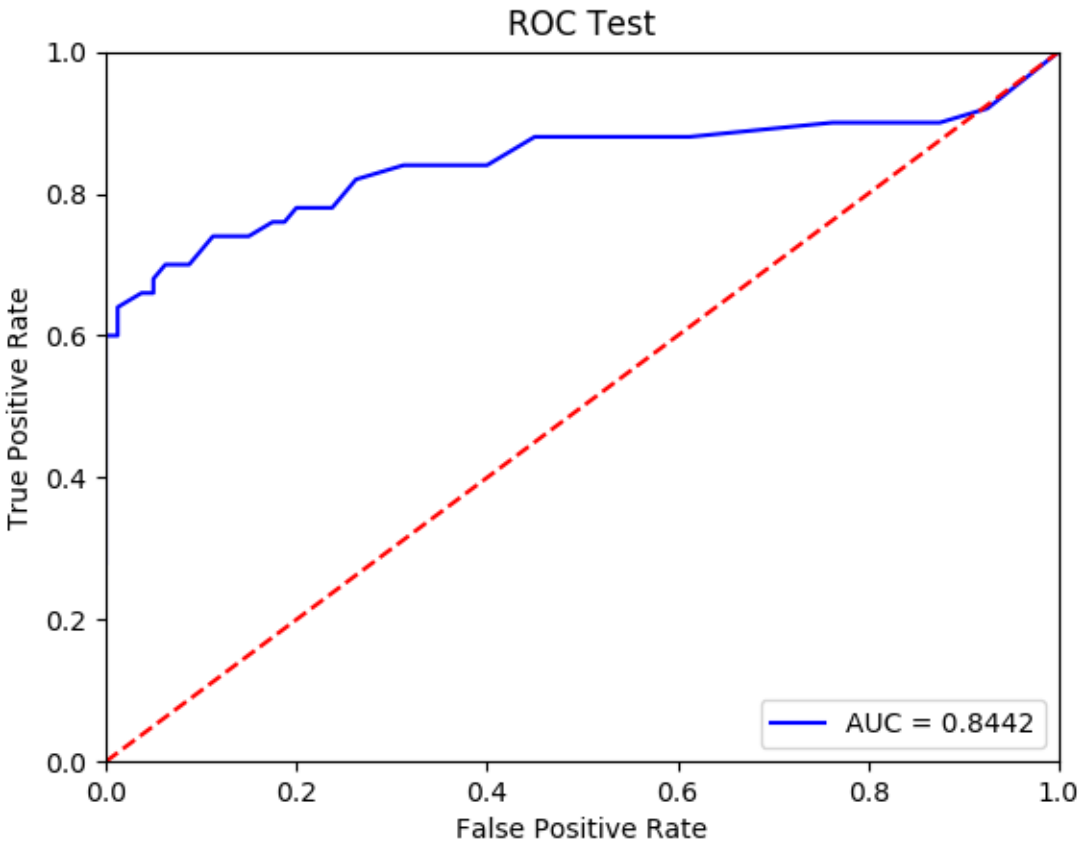
- Software Requirements:
  - 1. OS : Ubuntu 16.04
  - 2. Programming Languages: Python 3.5
  - 3. Deep Learning libraries : Tensorflow (v0.12.1)
  - 4. Support libraries: OpenSlide, SciKit, NumPy

Experiments: METRICS

- ROC:
  - Helps to measure classifier accuracy.
  - The ROC curve helps to create detailed sensitivity (true positive rate) vs 1-specificity (false positive rate) report

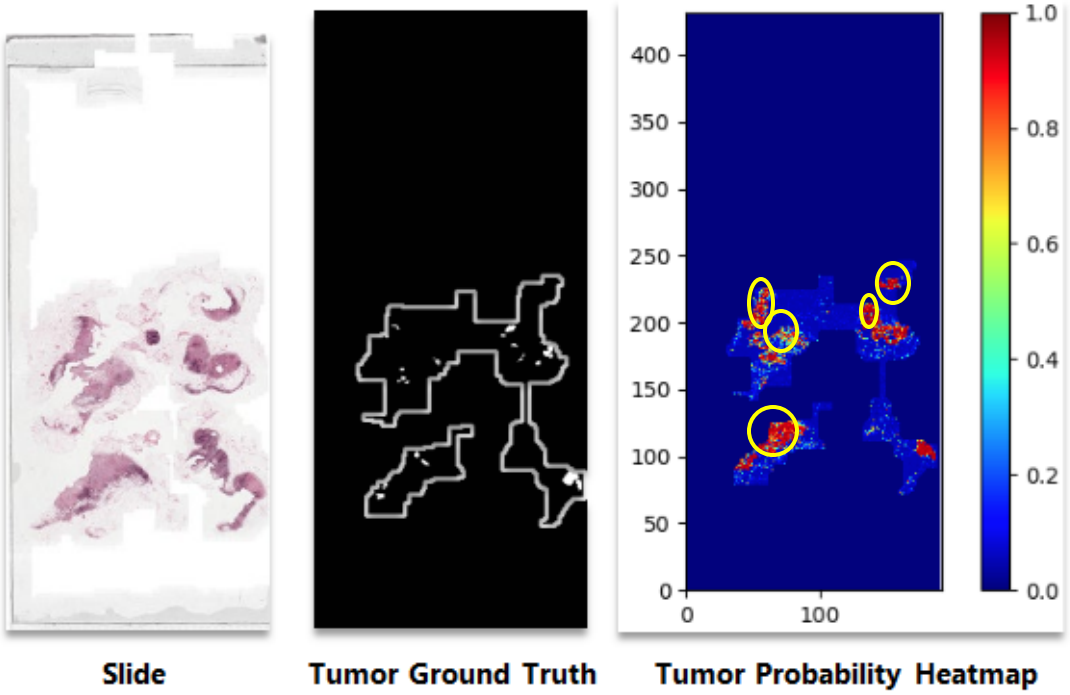
Experiments: MODEL D-1

- ROC for Deep model D-1:



Experiments: MODEL D-1 PROBLEMS

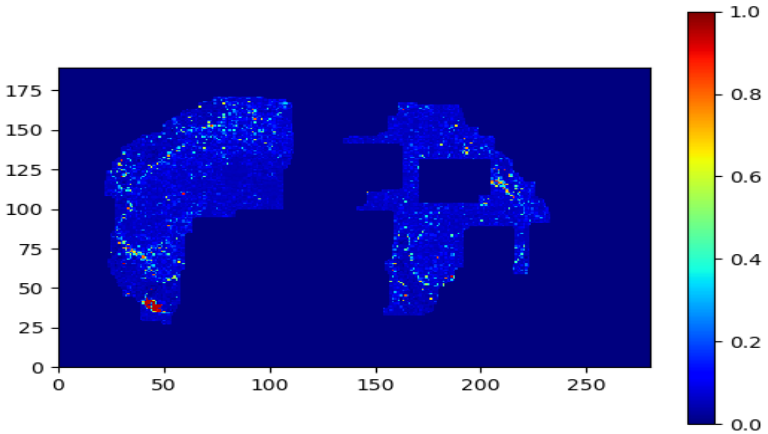
- Low AUC: 84.42%
  - Produces lots of false positives
  - **Reason:** in-comprehensive training data - hard negatives patches from histological mimics of cancer were missing in training dataset



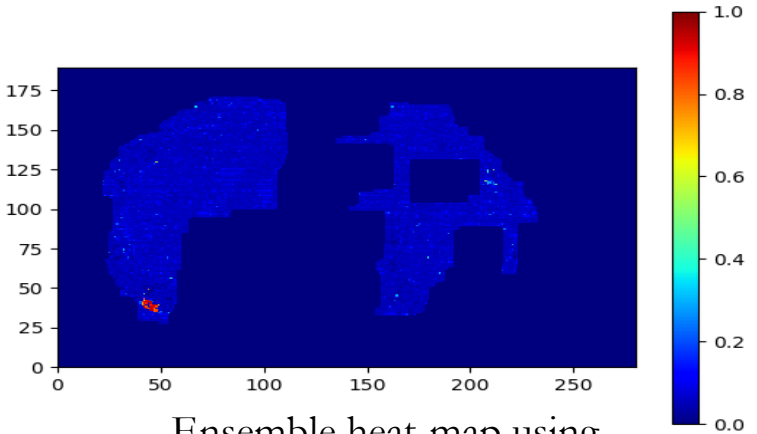
Experiments: ENSEMBLE METHOD

- Remove model D-1 false positives:
  - extract additional **~100k** hard negative patches corresponds to false positives of model D-l heat-maps
  - train deep model **D-2** with this enriched training dataset
  - using model D-2, build heat-map for each WSI
  - Build **ensemble** heat-maps by removing model D-1 false positives with model D-2 predictions

$$\text{pr(Ensemble)} = \begin{cases} \text{pr(D-2)}, & \text{pr(D-l)} >= 0.90 \ \& \ \text{pr(D-2)} < 0.50 \\ \text{pr(D-l)}, & \text{otherwise} \end{cases}$$



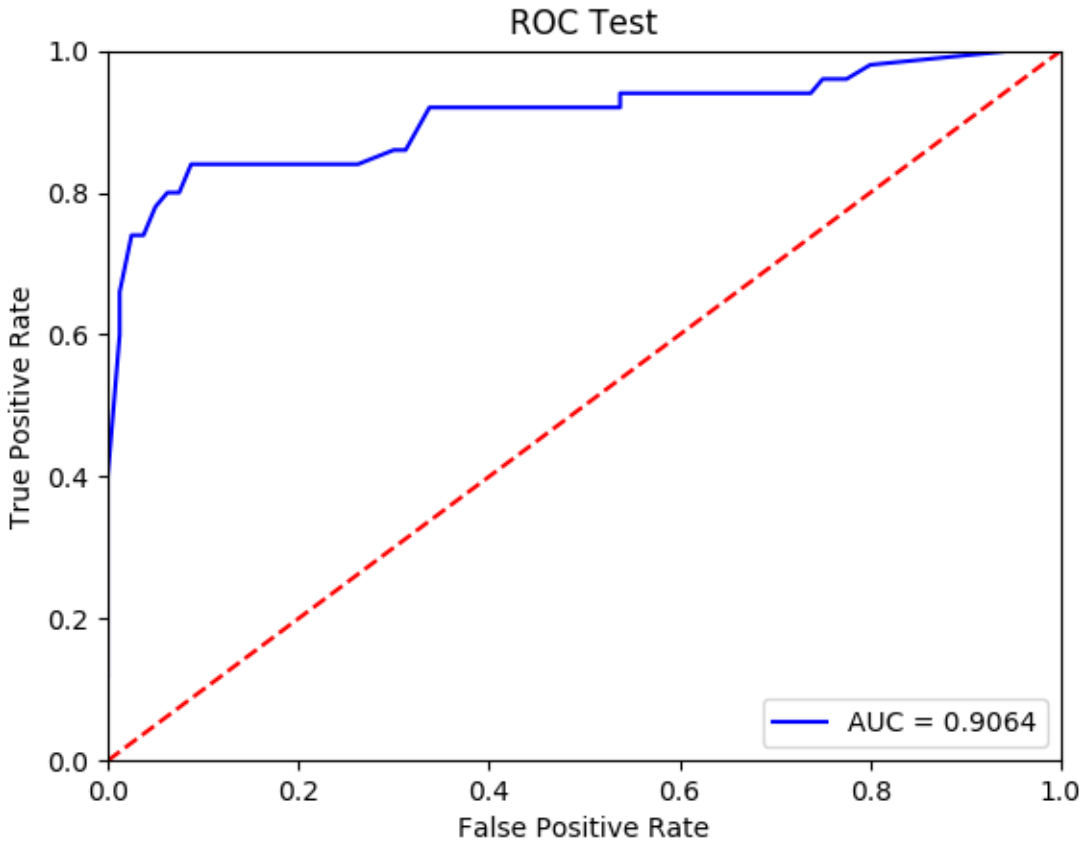
model D-1 heat-map



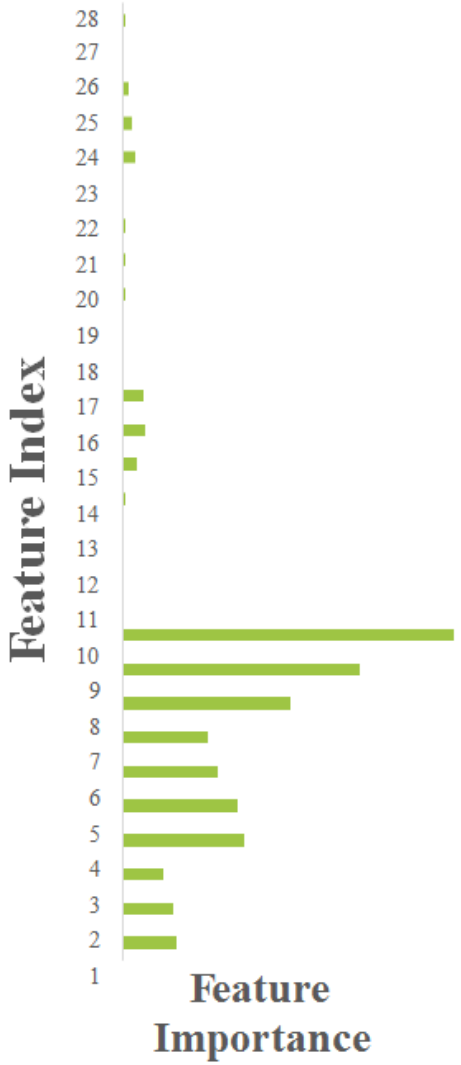
Ensemble heat-map using both models

Experiments: ENSEMBLE RESULT

- ROC for ensemble method:



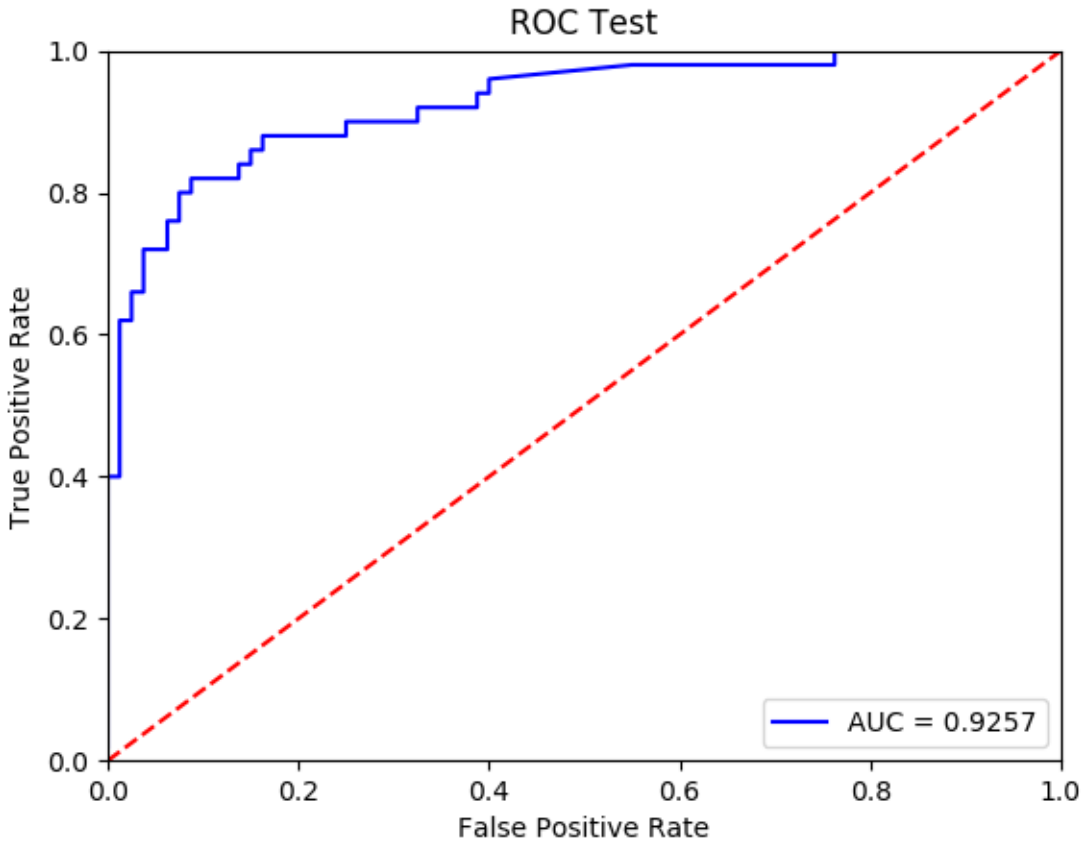
Experiments: FURTHER IMPROVEMENT



- Study feature importance map
  - remove features with low importance
  - remove undistinguished features (low correlation)
  - Removed **6** features
  - Keep **22** features
- Use better classifier - Support Vector Machine (SVM) instead of Random Forest

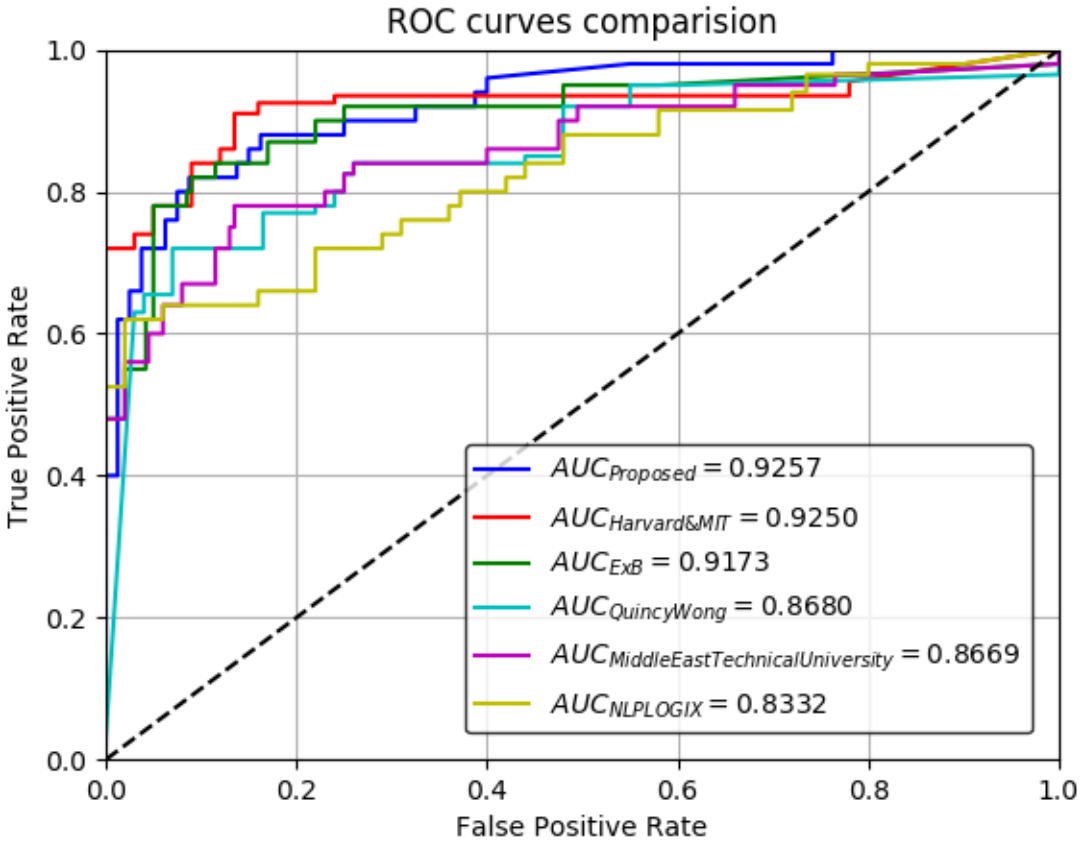
Experiments: FINAL RESULT

- Final ROC :



Experiments: RESULT COMPARISION

- ROC : comparison with Camelyon'16 **Top-5** methods



CONCLUSION & FUTURE WORK

- Conclusion
  - Developed deep-learning based classification pipeline for identifying metastatic breast cancer from histopathology images
  - Key aspects of our system includes enrichment of the training set with patches from regions of normal lymph node that the system was initially mis-classifying as cancer; use of a state-of-the art deep learning model architecture， and careful design of post-processing methods for the slide-based classification
- Future Work
  - Test proposed system on other large scale cancer datasets
  - Integrate staining normalization into proposed classification pipeline to eliminate variability induced by different staining techniques

Acknowledgements

- Supervising Professor
  - Dr. Junzhou Huang
- Thesis Committee
  - Dr. Heng Huang
  - Dr. Jia Rao
- Constant Support & Motivation
  - Ph.D. Mentor : Mr. Sheng Wang
  - Special Mention: Mr. Zhifei Deng, Mr. Jiawen Yao, Mr. Ashwin Raju
  - SMILE Lab colleagues
  - Friends & Family

Thank You.

Q & A