Medical Image Processing Course Project

Cancer Clinical Decision Support Enabled by Medical Image Processing of Histopathological Images from Cancer Patients

Introduction

The goal of this project is to assist students in learning the complete translational biomedical image processing pipeline for clinical decision support. Specifically, the students will work in teams to apply image processing and data mining techniques to cancer histopathological images, and to develop an objective and reproducible decision support for diagnosis and prognosis of cancers. The cancer images from The Cancer Genome Atlas (TCGA) will be provided to student to accomplish this goal. TCGA is a joint project by the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI). The goal of the TCGA project is to accelerate the understanding of cancer in order to more effectively diagnose, cure, and prevent cancer. TCGA provides high-quality genomic, proteomic, and imaging data and students will need to sign Data Access Requirement Form before giving access to these data. The TCGA cancer imaging data include whole slide tissue biopsy samples and MRI data for multiple types of cancer, where more than 1000 whole-slide images (~20,000×40,000-pixel in size) are provided for each cancer. Students will be provided sub-section images (512×512 pixels) that have been preprocessed by TA team after "quality control" procedure from few WSIs. The student teams will work on image segmentation, feature extraction, data mining, etc, which form the complete problem solving workflow in building a clinical decision support system.

Structure

This course term project is divided into three modules:

(1) Automatic segmentation of nuclear structures
(2) Image feature extraction and exploration
(3) Image classification

The modules are sequential blocks of a translational image processing pipeline for clinical decision making, and as such, methods developed in a module will be used by subsequent modules. Useful references are provided in the module description for students to start work with. Students will have totally ***nine calendar weeks*** to finish the entire project.

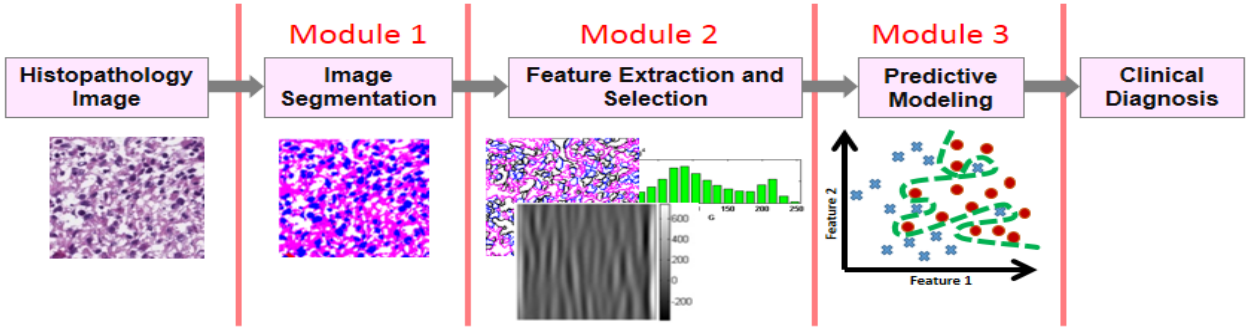


Fig. 1. Overview of the medical-image-processing-based clinical decision support systems

Evaluation Rubric

The project will be evaluated based on the following:

(1) Implementation
Teams are required to develop self-written codes in this project so to be able to fully understand the image analysis algorithms and parameters of these algorithms. If the team decides to use built-in MATLAB functions, the team needs to make explicit explanation to show its understanding of the methods, and the justification on why certain parameters ranges are used.

Required deliverables for grading:

a. Your submission must include a GUI wrapper to run your code.
b. You need to comment your codes to illustrate the function of each part. If your codes do not work for "test images" used by TA team during grading, you codes may receive partial credits if they are well documented.
c. You need to provide a precise README file with a manifest of functions and instructions to run the code. The file needs to be **explicit and concise** with no more than 500 words.

(2) Presentation – one presentation per module.

The Module-3 will be an in-class presentation for sharing and comparison. On Module-1 and Module-2, a one-to-one feedback session will be provided for each team by teaching team. *These will be scheduled around 3rd week and 6th week after the projects are assigned.*
Suggestions for presentation slides:

a. Problem statement (10%)
b. Methods (35%)
c. Results and discussion (35%)
d. Conclusion (10%)
e. References (10%)

(3) Report
Reports should contain the following sections:

a. Clinical need
b. Problem statement in technical terms (i.e. translating clinical need into image analytics tasks)
c. Data description (i.e. to illustrating your in-depth understanding of pathology)
d. Literature critique (i.e. every team member needs to **critique** at least 5 peer-reviewed conference papers or journal articles to answer "1H&5 W" of each paper. The contribution should be presented as CRITIQUE, and NOT just a copy and paste of abstract).
e. Design assumptions (i.e. to provide WHY you choose some algorithms over the others, and why you choose certain parameter range for the algorithm),
f. Methods (i.e. you need to provide concise mathematical descriptions, and define all variable properties – the size of the vectors, the dynamic range of the parameters etc.)
g. Implementation/Algorithms (i.e. you should provide good documentation of your codes in MATLab codes, and provide a clear FLOWCHART block diagram to show how all subroutines are calling/being called by others)
h. Answers to all the questions for the module (i.e. respond to module-specific questions).

i. Results/Discussion (i.e. you not only need to provide results in TABLEs or FIGUREs format, but also, you need to explain the meaning of the results, and to discuss whether these results are in-line with your prediction. WHY or WHY NOT?!)
j. References (i.e. again, if your team has N members, your references need to be 5*N to ensure every team member has contributed at least 5 references).

Teams will not get full credit if they simply report experiments and results with no discussion.

Key Dates: 9 Weeks for Whole Three Modules.

1. Project Period: Feb. 15th, 2016 (Monday) – April 18, 2016 (Monday)

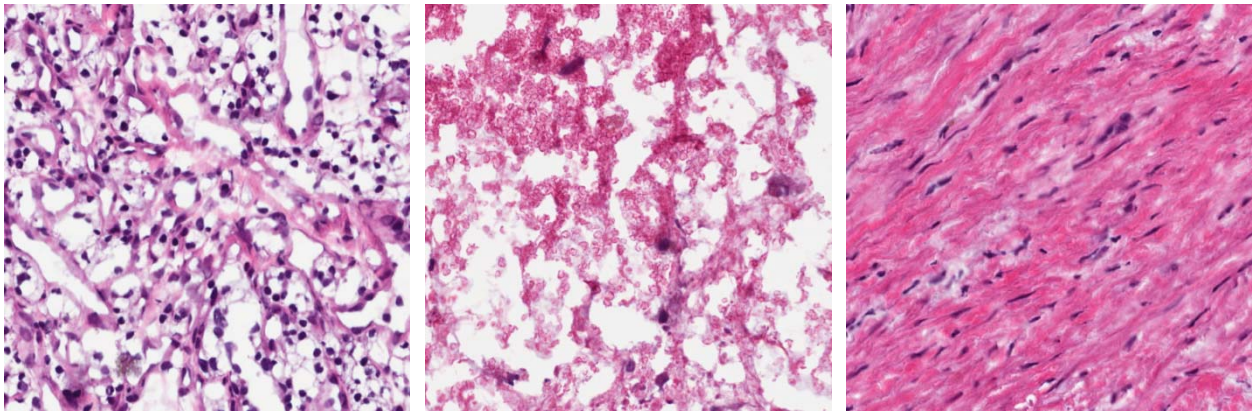
2. Feedback/Evaluation Milestone: 3rd week and 6th week of the project period. The written report is the lecture date after the oral feedback session.

Project Module 1: Image Segmentation

The goal of this module is to design, implement, and test two categories of color segmentation algorithm that can segment nuclear structures in histopathological images.

Data

1. For Undergraduate Teams: 300 digital microscopic images of hematoxylin and eosin (H&E) stained tissue sections of ***kidney clear cell carcinoma*** consisting of 100 tumor, 100 necrosis, and 100 stroma sections will be provided. H&E staining enhances three colors: blue-purple, white, and pink. These colors correspond to specific cellular structures.



Tumor Necrosis Stroma

Fig. 2. Example images in the kidney data set

Most researchers segment nuclear structures based on color properties. Thus, each image needs to be segmented into three parts:
1) Basophilic structures containing nucleic acids—ribosome and nuclei—tend to stain blue-purple;
2) Eosinophilic intra- and extracellular proteins in cytoplasmic regions tend to stain bright pink;
3) Empty spaces—the lumen of glands—do not stain and tend to be white.

2. For Graduate Teams: multiple sets of cancers will be provided:

(1) The first dataset is the same as the one for undergrad teams.

(2) The second dataset consists of 512×512-pixel rectangular portions of 100 whole-slide images (WSIs) of kidney clear cell carcinoma patients. Each patient is represented by 16 adjacent portions labeled as [patient name]_Tile_[row number]_[column number].png. Each patient is associated with the grade of cancer and the number of survival days.

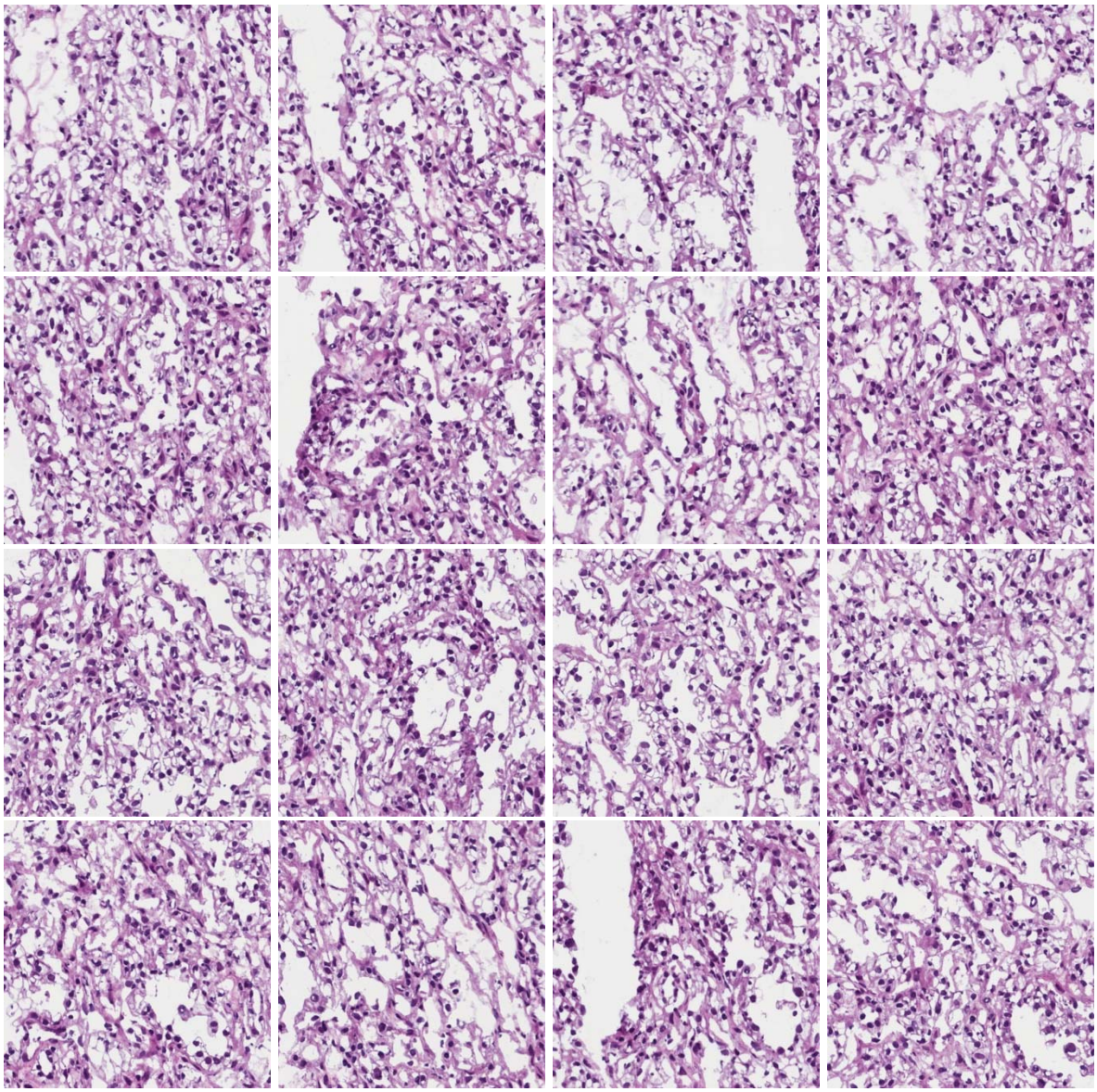


Fig. 3. Example images of the second kidney cancer dataset (1 patient with 16 adjacent images)

(3) The third dataset comprises of 512×512-pixel rectangular portions of 59 whole-slide images (WSIs) of pancreatic cancer patients. Each patient is represented by 16 randomly selected portions labeled as [patient name]_Tile.png. Each patient is associated with the grade of cancer and the number of survival days. However, in this module, there is no need to consider the clinical labels of the images when developing segmentation methods.

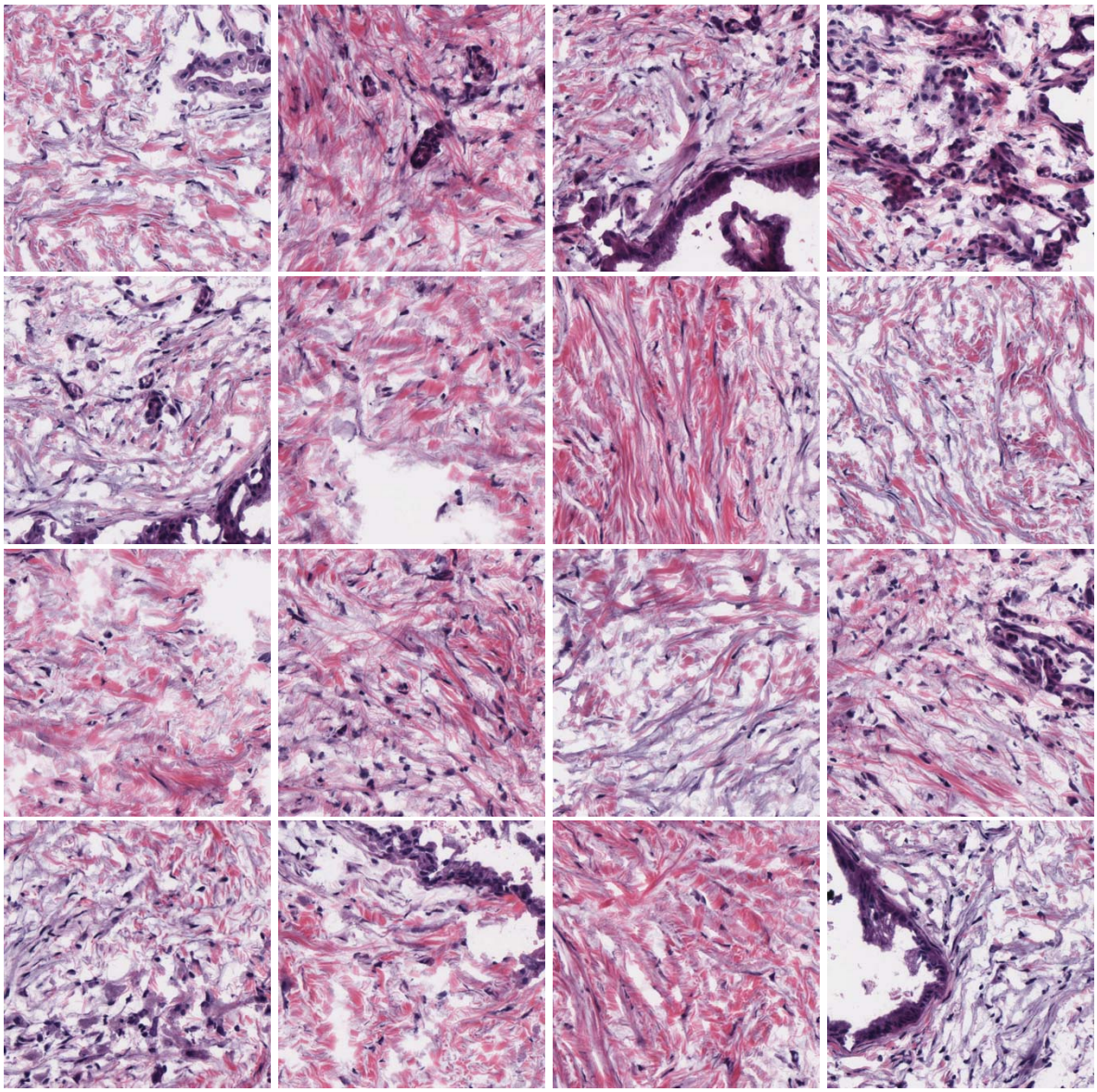


Fig. 4. Example images of the pancreatic cancer dataset (1 patient with 16 randomly selected images)

Your team is required to compare image segmentation methods applicable to all three datasets.

Deliverables

For completion of this module, teams will perform the following tasks:

(1) Survey, critique, design, and develop at least two categories of color-segmentation methods. Example references are provided, but the team still needs to do self-survey to find more possibilities close to what you are comfortable.
   a. Unsupervised method [1-3]

b. Supervised method [4-6]

Teams are required to survey and critique methods as required by overall project requirement (average 5 papers per team member). Then you can choose to implement any methods as required.

Tips: Because images provided for this module were acquired during different microscope setups, color normalization [7] or transformation may be helpful. Please cite appropriate references for your work and submit the PDF of referred papers with your submission.

(2) Implement at least two other metrics for comparing segmentation performance [8, 9]. Which performance metric is better and why?

(3) What is the effect of parameters, such as thresholds, distribution, and number of reference images (for supervised methods), on the segmentation performance?

(4) Discuss challenges and benefits of unsupervised vs. supervised methods. Which method performs better and why?

(5) Develop a GUI wrapper to run your code with the following:

a. Four image axes to display an input RGB, gray-level label (where nuclei, cytoplasm, and background are one, two, and three, respectively), pseudo-colored (where nuclei, cytoplasm, and background are blue, pink, and white, respectively), performance evaluation (depends on the evaluation metrics you choose in step 2). One example output is in the following Fig. 5.
b. Drop-down menus to select an image, method, and parameters.
c. A text panel that displays various performance metrics

(6) Write a concise and comprehensive report following the guidelines for report evaluation in term project guideline.

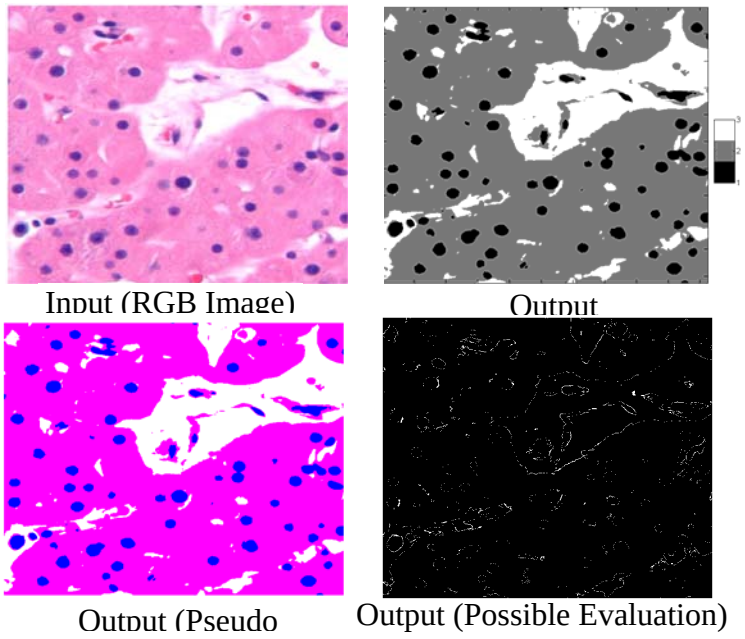


Fig. 5. Example output of the image segmentation result

Data Access

T-square: Resources-Project Data Sets folder

Notes: Dataset 2 and dataset 3 are split into several parts due to size limitation of T-square. Please make sure you download all parts and put them in the same folder before decompressing the files.

(A) Module-1 oral presentation of ALL 13 teams: Mar. 6th, 2016.

(B) **Module-1 written report: 11:59pm, Mar. 9th, 2016 (Wednesday). Due date for Module 1 final written report and MATLAB code after revision based on the feedback from presentation.**

References:

[1] M. Datar, D. Padfield, and H. Cline, "Color and texture based segmentation of molecular pathology images usING HSOMS," *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 292-295, 2008.

[2] K. Jun, H. Shimada, K. Boyer, J. Saltz, and M. Gurcan, "Image analysis for automated assessment of grade of neuroblastic differentiation," *Biomedical Imaging: From Nano to Macro, 2007 ISBI 2007 4th IEEE International Symposium on*, pp. 61-64, 2007.

[3] O. Sertel, J. Kong, U. Catalyurek, G. Lozanski, J. Saltz, and M. Gurcan, "Histopathological Image Analysis Using Model-Based Intermediate Representations and Color Texture: Follicular Lymphoma Grading," *Journal of Signal Processing Systems*, vol. 55, pp. 169-183, 2009.

[4] S. Kothari, J. H. Phan, R. A. Moffitt, T. H. Stokes, S. E. Hassberger, Q. Chaudry, *et al.*, "Automatic batch-invariant color segmentation of histological cancer images," *Conf Proc IEEE Int Symp Biomed Imaging, ISBI*, pp. 657-660, 2011.

[5] K. Z. Mao, P. Zhao, and P. H. Tan, "Supervised learning-based cell image segmentation for p53 immunohistochemistry," *Biomedical Engineering, IEEE Transactions on*, vol. 53, pp. 1153-1163, 2006.

[6] C. Meurie, G. Lebrun, O. Lezoray, and A. Elmoataz, "A comparison of supervised pixels-based color image segmentation methods. application in cancerology," *WSEAS transactions on Computers*, vol. 2, pp. 739-744, 2003.

[7] D. Magee, D. Treanor, D. Crellin, M. Shires, K. Smith, K. Mohee, *et al.*, "Colour Normalisation in Digital Histopathology Images," *Proc. Optical Tissue Image analysis in Microscopy, Histopathology and Endoscopy (MICCAI Workshop)*, pp. 100-111, 2009.

[8] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "Toward Objective Evaluation of Image Segmentation Algorithms," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, pp. 929-944, 2007.

[9] H. Zhang, E. Fritts Jason, and A. Goldman Sally, "Image segmentation evaluation: A survey of unsupervised methods," *Computer Vision and Image Understanding*, vol. 110, pp. 260-280, 2008.

Project Module 2: Feature Extraction and Selection

The goal of this module is to extract, select, and explore image features in histopathological images. To gain insight about feature selection process, teams will develop and compare different feature extract, feature ranking, and feature reduction methods.

Data

The same datasets for module 1 will be used in this module.

Deliverables

For completion of this module, teams will perform following tasks:

(1) Conduct literature survey on histopathological image analysis (at least 5 peer-reviewed papers). Summarize **at least 5 types of features** reported for histopathological images, and **compare at least 5 feature extraction methods** that can be used to extract features to show your understanding of the strength or limitation of each method from the paper.

(2) Among many possible feature extraction methods, implement a few methods to extract three categories of image features (more than 100 features in total) as reported in [10-12]:
   a. Color
   b. Morphology such as shape and topology (after segmenting stains based on methods developed in Module 1)
   c. Texture such as wavelet, GLCM, and fractal

(3) Implement/Apply at least **two feature ranking methods** using supervised statistics such as mutual information, correlation, and ANOVA. Usually, the list of top few features selected using different statistical methods will be different. Please analyze the difference and then provide rationale on why they differ.

(4) Implement/Apply at least **two methods** for data dimensionality reduction such as PCA, LDA, graph embedding, etc. Discuss the loss of information when samples are represented by reduced data dimension in the transformed domain.

(5) Develop at least one simple visualization such as the boxplot, scatterplot, etc. to show the separation of classes in image feature space.

(6) Please use the GUI wrapper you developed in Module-01 to run your code **with the following:**
   a. Adding a drop-down menus to select, individual AND combinations of feature selection methods, feature transformation method, visualization, and parameters.
   b. A text panel that displays various performance metrics

(7) Write a concise, comprehensive report following the guidelines for report evaluation.
   a. Please include detailed description of ***feature extraction, feature ranking, and feature reduction*** methods that you surveyed and implemented.

b. Please compare different feature extraction, feature ranking, and feature reduction methods you implemented to highlight the difference, the strength, and the limitation of each method.

Key Dates

A. **Module-2 oral presentation of ALL teams: Mar. 28th, 2016 (Monday).**

B. **Module-2 written report: 11:59pm, Mar. 30th, 2016 (Wednesday) Due date for Module 2 final written report and MATLAB code after revision based on the feedback from presentation.**

References

[10] M. N. Gurcan, L. Boucheron, A. Can, A. Madabhushi, N. Rajpoot, and B. Yener, "Histopathological Image Analysis: A Review," *IEEE Rev Biomed Eng*, vol. 2, pp. 147-171, 2009.

[11] S. Kothari, J. H. Phan, A. N. Young, and M. D. Wang, "Histological image feature mining reveals emergent diagnostic properties for renal cancer," *Conf Proc IEEE Bioinform Biomed, BIBM*, 2011.

[12] L. Boucheron, "Object-and spatial-level quantitative analysis of multispectral histopathology images for detection and characterization of cancer," PhD thesis, University of California, Santa Barbara, 2008.

Project Module 3: Prediction Modeling

The goal of this module is to develop and validate computer-based prediction models as presented in papers [13, 14]. This is the foundation for clinical diagnosis decision support system development based on cancer biopsy images. Teams will build prediction models using different classifiers; apply different cross-validation (or internal validation) schemes and different performance metrics on training dataset to train the model; and validate the predictive models using test dataset (i.e., external validation).

Data Access T-square: Resources-Project Data Sets folder

The same datasets for module 1 will be used in this module. But patient class labels will be provided.

Notes: Dataset 2 and dataset 3 are split into several parts due to size limitation of T-square. Please make sure you download all parts and put them in the same folder before decompressing the files.

1. For Undergraduate Teams: 300 digital microscopic images of hematoxylin and eosin (H&E) stained tissue sections of ***kidney clear cell carcinoma*** consisting of 100 tumor, 100 necrosis, and 100 stroma sections will be provided. Teams will develop a multiclass classification model for this dataset. Please use 60 images from each class, indexed 1 to 60, as the training set and 40 images from each class, indexed 61 to 100, as the testing set.

2. For Graduate Teams: three datasets of cancers will be provided:

1) The first dataset is the same as the one for undergraduate teams.

2) The second set comprises of 512×512-pixel rectangular portions of 100 whole-slide images (WSIs) of kidney clear cell carcinoma patients. Each patient is represented by 16 adjacent portions labeled as [patient name]_Tile_[row number]_[column number].png.
   Out of these 100 WSIs, 60 images are in the training set while the other 40 are in the testing set (refer to the spreadsheet). Each patient is associated with the grade of cancer and the number of survival days. Please refer to the spreadsheet for clinical labels. For this dataset, teams will develop two binary models: high vs. low grade as well as high vs. low survival.

3) The third dataset comprises of 512×512-pixel rectangular portions of 59 whole-slide images (WSIs) of pancreatic cancer patients. Each patient is represented by 16 randomly selected portions labeled as [patient name]_Tile.png. Each patient is associated with the grade of cancer and the number of survival days.
   Out of these 59 WSIs, 34 images are in the training set while the other 25 are in the testing set (refer to the spreadsheet). Each patient is associated with the grade of cancer and the number of survival days. Please refer to the spreadsheet for clinical labels. For this dataset, teams will develop two binary models: high vs. low grade as well as high vs. low survival.

Note: The second and third datasets have multiple images for a patient. **Teams may combine image features from all portions of a patient's WSI** to represent the patient and predict patient's class, or predict class for each portion and combine predictions to predict patient's class [15]. In either case, all the portions belonging to a patient can be either in training or testing set.

Deliverables

For completion of this module, teams will perform the following tasks:

(1) Conduct literature survey on computer-based classifiers for image analysis **(at least 10 peer-reviewed papers)** to identify both **linear classifiers and non-linear classifiers**. Then summarize **at least five typical classifiers for decision making**. Examples include Fisher discriminant analysis) [16], support vector machine (SVM) [17] , neural networks (NN) [18], k-nearest neighbor (KNN) [19], and decision tree [20] to show your understanding of the strength or limitation of each method.

(2) Develop prediction models using **three classifiers** out of the 5 classifiers you have surveyed and critiqued: Fisher discriminant analysis (MATLAB classify function), SVM, NN, KNN, and decision tree.
Try to experience different classifiers and optimize classifier parameters for **all three prediction model building for the following 3 clinical endpoints for kidney cancer datasets:**
   a. Tumor vs. Necrosis vs. Stroma
   b. High vs. Low kidney grade
   c. High vs. Low kidney survival

(3) Develop **two cross-validation schemes** (or internal validation) by training the classifiers using the training dataset
   a. N-iterations of m-fold cross-validation
   b. Leave-one-out cross-validation or bootstrapping for classifier optimization [13, 14]

(4) Develop **four performance metrics to evaluate** your predictive model performance when applying to test dataset (i.e., external validation)
   a. Area Under the Curve (AUC)
   b. Matthews Correlation Coefficient (MCC)
   c. F-score
   d. Accuracy [21]

NOTE:

For Undergrad teams, you only need to do kidney cancer to satisfy minimum requirement.

For Graduate teams, you can optimize the classifier parameters for pancreatic cancer dataset in a similar manner. For example, you can explore questions such as (i) what are cancer-specific image features versus cancer-non-specific features; (ii) which predictive models work best for which cancer; and (iii) which performance metric works better for which cancer.)

(5) Write a concise and comprehensive report on the cancer-biopsy-based clinical diagnosis decision making system following the guidelines below for report evaluation.

a. Please include detailed description of **performance metrics, cross-validation schemes, optimization parameters for different classifier selection** in predictive model building
b. Please **compare different metrics** of all your classifiers and predictive models in validation using the external test dataset.
c. Please **provide discussion** about what is the strength and limitation of each classifier in predictive modeling building
d. Please **provide discussion** about what causes the difference among different predictive models using different performance metric; and which metric is more appropriate for your problem at study.
e. Please provide a direction for future improvement.

NOTE:

1) After this module, you will be asked to write and submit **self-evaluation and peer-evaluation** to explicitly state the contribution of each team members in all aspects of this project throughout the semester (if you changed the team in the middle of the semester, just note it).

2) Because of the critical need for the clinical decision support system in the U.S. and around the world, any team who has proposed innovative ideas and has shown improvement in this final phase of the project will be continually supported for future presentation to the wider technical and clinical community in the form of technical conference pursuit.

Key Dates

A. Module-3 oral presentation of most teams on Apr. 18th (Mon.) 1:00pm-3:00pm, and a few teams on Apr. 20th (Wed.) 1:00pm-1:30pm.

(4/20/2016: 1:30pm-2:30pm for Final Exam Review, and 4/25/2016 will be Reading Period.)

B. Module-3 written report: 11:59pm, Apr. 22th, 2016 (Fri.): Due date for Module 3 final written report and MATLAB code after revision based on the feedback from presentation.

References

[13] R. Bellazzi and B. Zupan, "Predictive data mining in clinical medicine: current issues and guidelines," *Int J Med Inform*, vol. 77, pp. 81-97, 2008.

[14] F. Pereira, T. Mitchell, and M. Botvinick, "Machine learning classifiers and fMRI: a tutorial overview," *Neuroimage*, vol. 45, p. S199, 2009.

[15] J. Kong, O. Sertel, H. Shimada, K. L. Boyer, J. H. Saltz, and M. N. Gurcan, "Computer-aided evaluation of neuroblastoma on whole-slide histology images: Classifying grade of neuroblastic differentiation," *Pattern Recognition*, vol. 42, pp. 1080-1092, 2009.

[16] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. Mullers, "Fisher discriminant analysis with kernels," in *Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop.*, 1999, pp. 41-48.

[17] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, p. 27, 2011.

[18] A. K. Jain, J. Mao, and K. M. Mohiuddin, "Artificial neural networks: A tutorial," *IEEE computer*, vol. 29, pp. 31-44, 1996.

[19] T. M. Mitchell, "Machine learning. 1997," *Burr Ridge, IL: McGraw Hill*, 1997.

[20] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 21, pp. 660-674, 1991.

[21] S. García, A. Fernández, J. Luengo, and F. Herrera, "A study of statistical techniques and performance measures for genetics-based machine learning: accuracy and interpretability," *Soft Computing-A Fusion of Foundations, Methodologies and Applications*, vol. 13, pp. 959-977, 2009.